



A Corpus-Based Study on Language Features and Literary Themes in the Yellow Wall-Paper and Herland by Charlotte Perkins Gilman

Hui-Chuan Lu

(National Cheng Kung University)

Kai-Ling Liu

(National Cheng Kung University)

Chien-Ting Yeh

(National Cheng Kung University)

Ya-Jie Chen

(National Cheng Kung University)

Lu, H. C., Liu, K. L., Yeh, C. T., & Chen, Y. J. (2022). A corpus-based study on language features and literary themes in the Yellow Wall-Paper and Herland by Charlotte Perkins Gilman. *Asia Pacific Journal of Corpus Research*, 3(1), 21-34.

This study aims to apply corpus-based approach to analyze *The Yellow Wall-Paper* and *Herland* written by Charlotte Perkins Gilman, a women's rights activist in the late nineteenth-century America. Although both works have attracted feminists' attention to the woman question that concerned Gilman, discussion on her language features and their relation to the literary themes of these two works is still in need. In this corpus-based analysis, we argue that the main themes of different literary works can be revealed through linguistic patterns identified by number and gender features of nouns and pronouns in the contrast of two works and a balanced corpus. The linguistic features (number and gender) have been related with two themes, the 'group and individual' and the 'feminine and masculine', and are further interpreted in terms of mothering and feminine consciousness. By adopting linguistic approach, our study provides quantitative and qualitative evidence to verify the established themes and arguments of these literary texts.

Keywords: Corpus-Based, Mothering, Gilman, Number, Gender

1. Introduction

Studies of linguistics and literature traditionally used to follow the conventions of their own disciplines. However, with the advancement of technology, disciplinary scope and new research insights can benefit researchers of a cross-disciplinary team. With this belief, the first (a linguistic major) and the second (a literature major) authors started this experimental tryout. Currently, the application of corpus-based approach to literary studies has primarily employed corpus tools to explore the linguistic regularities of literary texts. What triggered our collaborative study was a late nineteenth-century American woman writer, Charlotte Perkins Gilman (1860-1935). A women's rights activist and prolific writer, Gilman always felt language per se was hardly at her command. For her, language was "a rational tool that could be used to explore and mold ideas, but one that was inadequate before the chaos of emotion and psychological pain that tortured her throughout her life" (Weatherford, 1999, p. 58). Among her many publications, *The Yellow Wall-Paper* (1892, TYWP, henceforth) and *Herland* (1915) are still widely read today. The former is an autobiographical short story, written based on her own postpartum depression after she gave birth to her only daughter; the latter is a utopian novel, which describes an ideal all-women country where each woman works to

produce or re-produce according to her talent and condition, rather than by gender stereotypes as practiced in "Our Land." If we take TYWP as showcasing "the inadequate language [for] the chaos of emotion and psychological pain," and Herland as showcasing "a rational tool that could be used to explore and mold ideas," then, two research questions follow:

- (1) What are the linguistic features and patterns found in the two works?
- (2) How do the detected linguistic features and patterns pertain to the themes of the two works?

To answer these two questions, we propose to use the corpus-based approach to analyze language use in the two target works. In the following, we will first review pertinent literature, describe the research methodology, elaborate on data findings, analysis, and discuss. Finally, we will provide our conclusion to this study.

2. Literature Review

The corpus-based approach has been applied to various sub-fields of linguistics studies in the last two decades because of its advantages in presenting the objectivity of data (Leech, 1992). This approach then has been adopted for other humanity studies such as the earlier corpus-based literature analysis published by Ross in 1973. Especially, stylistics study has been developed as a new stream that integrates linguistic approach to literature analysis. According to Mahlberg (2010), "researchers have started to become interested in works that can be called 'corpus stylistics,' which employs methods and approaches of corpus linguistics and links them with concerns in literary stylistics and literary criticism" (p. 295). Furthermore, Hardy (2007) indicates that computational stylistics is valuable because "it uses the computational power and convenience of the personal computer to search, research, calculate, and re-calculate linguistic patterns of interest in literary works" and "ideally the analysis produces insight into the remarkably finely detailed formal level at which some literary meaning is produced" (p. 24). As a result, the searching and calculating capacities of computers enables scholars to search texts much more efficiently. For example, Sveen (2005) applied corpus-based approach to compare adjectival descriptions of characters in Victorian and contemporary British children's fiction. With the help of "computer-readable texts, and tools", she pioneered "to explore systematically a large amount of material" and discuss a diachronic comparison of adjectival descriptions of characters in British children's fiction (p. 17).

In addition to the benefit of convenient analytical tools, studying literature from a linguistic perspective by using quantitative methods may break two limitations of traditional literary studies. The first limitation is that while "[the] definition and the exact nature of literary criticism therefore changes in the course of time as different social and political conditions prevail" (Fischer-Starcke, 2010), that of linguistics stays more unaffected. The other limitation is that "[linguistic] features that are not prominent in the text are frequently not recognized and are therefore only rarely analyzed in literary studies" (Fischer-Starcke, 2010). Studies by Mahlberg (2010) and by Hardy (2007) stand as examples of breaking through these two limitations. Mahlberg analyzes the keyword, 'civility', in *Pride and Prejudice* and compares it through the reference corpus. As a result, he argues that the novel's theme is "misunderstandings and misjudgments that result from a mismatch between outward civilities and true virtues of characters" (p. 296). Moreover, Hardy compares Flannery O'Connor's work with the Brown fiction sub-corpora, aiming to introduce and explore words related to body parts in O'Connor's fictions. In order to reach the goal, Hardy uses various procedures, including frequency counts, keyword indexing, collocations, contextualized searches with the internet-based program Textant (Hardy, 2007). He points out that collocational analysis should include at least "a two-step process", and "one must determine what, if any, grammatical patterns appear to be represented among the collocations and then

what, if any, literary significance might be attached to those patterns” (p. 145). As the review above indicates, the key point to connect linguistic and literature perspectives falls on the common factors of these two fields that can be analyzed through the assistance of corpus tools to reflect literature features. According to Mahlberg, “the application of corpus techniques to the study of literary texts has to combine quantitative and qualitative analysis to provide useful insights” so that “literary scholars may profit from corpus methods, and through the engagement with literary criticism, corpus linguists may be able to develop more specific corpus resources and methods” (Mahlberg, 2010, p. 298).

Charlotte Perkins Gilman’s life is a worth reading story. She and her younger brother were raised by their mother after her father abandoned them. The family moved a lot and was always in poverty, since her mother and Gilman herself when she grew older, little access to paid work. In her late teens, she started to think over “the injustices under which women suffered, and still more with the ill effects upon all mankind of this injustice” (Gilman, 1935, p. 61). In 1885, after she gave birth to a daughter, postpartum depression caused her mental breakdown and she found herself being “trapped by the role assigned to the wife within the conventional nineteenth-century marriage” (Hedges, 1973, p. 46). TYWP was published in 1892 not only as a “witness to the personal and social anguish of its author,” but also as “an indictment of the incompetent medical advice she received” (Hedges, 1973, p. 46). A story which vividly depicts the narrator’s mental process under the popular treatment of depression at that time—the rest cure, TYWP, unlike other works of Gilman, reveals the emotional world of the protagonist which Gilman created based on her personal experiences (Lane, 1990). The outspoken way Gilman allows herself to dig out feelings and fears is even more obvious than in her autobiography, or letters. The focus of TYWP is “sexual politics,” (Dock, 1988, p. 19) which severely criticizes how the contemporary requirements of women to be selflessly caring wives and mothers drove women into insanity (Dock, 1988, p. 19; Schörkhuber, 2008). Herland, on the other hand, is a utopian fiction which pictures an ideal women society where gender ideology of Gilman’s contemporary society is “ridiculed” (Lane, 1990, p. 293). In contrast with TYWP, a short story full of tense emotions, Herland unveils rationality could break the gender dichotomy and push men and women in pursuit of “happiness” (Lane, 1990, p. 304). The idea of the communal motherhood and roles of women according to their capability are the major arguments of the fiction.

In the following section, we will explain the analysis of linguistic features in these two works and elaborate on the two research questions.

3. Methodology

Previously, we mentioned that we will use the corpus-based approach to analyze language use in the two target works. To be more specific, we focus the language use in the literary works on the so-called phi-features. Harbour et al. (2008) defined them “to be those involved in predicate-argument, typically person, number, and gender” (p. 2) and Kerstens (1993) mentioned that “the set of phi-features traditionally includes the pronominal features of person, number and gender” (p. 1). In our study, number feature of nouns, pronouns and verbs on one hand, and gender feature of nouns and pronouns on the other hand can be easily extracted by tools because they are searchable, i.e. they can be limited to certain amount of parts of speech (POS) and words. According to the search results, the combination of using corpus-based method to analyze phi-features in literary works has not been done in the previous studies. Hence, the present study is creative in providing this critical perspective. The Yellow Wall-paper consists of 6,171 words and Herland consists of 53,644 words. British National Corpus¹ (BNC, henceforth) is the reference corpus composed of 100,000,000 words. We use Mike Scott’s

¹ The British National Corpus, version 3 (BNC XML Edition) (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Available at: <http://www.natcorp.ox.ac.uk/>

WordSmith Tools (version 5)² to facilitate the analysis of TYWP, Herland and the reference corpus (BNC). The two literary texts are downloaded from the Electronic Text Center of University of Virginia Library. These text files undergo POS tagging through the Constituent Likelihood Automatic Word-tagging System (CLAWS4) POS tagger to facilitate the classification of grammatical categories. Then, we save both types of text files as different corpora, POS-tagged and plain versions for different purposes in the following analysis. The former type is saved for filtering same type of POS (for example, nouns and pronouns) while the latter one is for searching specific words such as ‘we, mother’. With respect to the two linguistic features mentioned in the previous section, the number feature can be observed from nouns, pronouns and verbs while the gender features are involved in the first two parts of speech.

There are three main functions in WordSmith Tools, WordList, KeyWords and Concord. Through the first function, a word list for each literary work can be obtained. By observing the frequency order of words in the lists, we will find high and low frequent words used in both works. In addition to comparing both lists extracted from the two texts studied, we also contrast their word lists with that of the reference corpus, the BNC wordlist downloaded from WordSmith web site. By comparing the three word lists, we will know the similarities and differences among them and the difference between usages in the two studied texts and those in a balanced corpus. Furthermore, with the second function of the auxiliary tool, we are able to acquire the keywords by importing any two word lists of compared corpora in order to contrast their significant differences. Through correlation statistics, a word which appears in unusual high frequency in one corpus but appears in low frequency in another corpus (positive correlation) or vice versa (negative correlation) will be listed to draw special attention. The high frequent words in a corpus of text might reflect its importance in a literary text since they appear repeatedly, i.e. they should be related to the themes of a work, especially when they are keywords in contrast with the balanced reference corpus. Following this, keywords of two compared texts can distinguish two texts and lead to distinct themes presented in different texts. Furthermore, it is possible to study specific single words or strings of words by using the third function of concord to search words in contexts to observe their collocations or clusters of adjacent sequence of various elements and analyze usage patterns in context for further and deeper analysis of each literary work. Going one step further, other related factors or collocated clusters can be discovered through association with the searched elements of interest. The results of searched elements or strings of collocation can be compared and contrasted to gain better understanding of the texts. Those previous procedures can derive word lists, keywords and collocations in a few seconds compared to arduous work of traditional manual research method. It is worthwhile to mention that in calculating the frequency, we do consider the percentage of appearance compared to the whole group taking into account the total words of each corpus.

Finally, in order to enrich the quantitative results and examine in detail words with higher tendency of usage from a semantic perspective, we conduct textual analysis on the sentences containing the specific word by annotating data with different classification of meanings. For example, we locate the word ‘mother’ and filter all the sentences containing this word. Then we analyze all those sentences to understand the word, ‘mother’ in a context to find their similarities and difference.

² Scott, M. (2008). *WordSmith Tools Version 5* [Computer Software]. Liverpool: Lexical Analysis Software. Available at: <http://www.lexically.net/wordsmith/>

4. Results and Discussion

4.1. The Number Feature

As shown in Table 1, by using the word list function of WordSmith Tools, we have acquired 10 words that appear most frequently in the three corpora searched, TYWP, Herland, and BNC.

Table 1. List of the Top 10 Words

Text	Top 10 Words	Word	Order	Frequency (%)
TYWP (6,171)	I, and, the, it, to, a, is, that, of, in	I	1	305(4.94%)
Herland (53,644)	the, and, of, to, a, we, in, that, I, was	We I	6 9	1,026(1.91%) 750 (1.40%)
BNC (100,000,000)	the, of, and, #, a, in, that, is, it, for	---	---	---

From Table 1, functional words ('the', 'a'), prepositions ('to', 'of', 'in'), and conjunction ('and') appear in the top 10 word lists. The remaining words on the top 10 word list are pronouns and verbs. However, while first-person pronouns ('we' and 'I') prevail in TYWP and Herland, the third-person singular neutral pronoun, 'it', prevails in BNC. This distinction of the number feature of pronouns between the two literary works and such a balanced corpus as BNC endows significance to the genres that Gilman employed for her works. The first-person narrator in TYWP tells her gradual regression into insanity. The story takes place in a distant forlorn countryside estate; even though the narrator is accompanied by her sister-in-law who is in full charge of her new-born baby and the household, and her medical doctor husband who visits her now and then, the narrator is in reality isolated by her husband who prescribes her a strict rest cure, cautioning her against any social and intellectual activities. The more the narrator is discouraged by her husband to 'reason' with him, the more she turns to her diary for interlocution; yet with the persistent silence of this interlocutor, the expected dialogue becomes the monologue. Curiously, when the narrator identifies with the woman behind the yellow wallpaper who eventually releases herself from the cell of the wallpaper and crawls along the room, she becomes the woman behind the wallpaper. Ironically, the release of the woman from the cell of the wallpaper indicates the release of the narrator from the confining social cultural prescriptions of women only by getting herself into insanity. Freedom for women is gained at the cost of exclusionary insanity. In this way, the frequent appearance of the 'I' in the story is a fact not only because this is an autobiographical narrative but because the paradoxical 'I' refers to more than the narrator but a woman, the woman, and women, whose silence, loneliness, and the feeling of not being rightfully appreciated makes TYWP the piece of work so emotionally embedded that Gilman had to complain language was not at her demand.

On the top 10 word list, there is one word that might become the counterexample for our hypothesis and needs to be accounted for. That is the verb 'was'. Since it is the singular form, we expect it to be associated with the subject of singular form according to the principle of subject-verb agreement. However, this will be considered as exceptional or conflict because if we follow our hypothesis and deduction based on previous observation related to the plural subject 'we' in considering plurality as a special linguistic feature in Herland. Before getting into further discussion, we will show the result of keywords (i.e. Table 2) derived from the two texts, which is significant since it undergoes the probability statistics of correlation ($P \leq 0.05$).

Table 2. Result of Keyness (TWYP vs. Herland)

Word	TYWP	Herland	Keyness	P
I	305 (4.94%)	750 (1.40%)	287.19	0.000
Is	131 (2.12%)	267 (0.50%)	150.41	0.000
Me	70 (1.13%)	156 (0.29%)	72.63	0.000
Am	23 (0.37%)	8 (0.01%)	70.89	0.000
It	173 (2.80%)	705 (1.31%)	69.23	0.000
My	40 (0.65%)	119 (0.22%)	28.36	0.000
We	18 (0.29%)	1,026 (1.91%)	-124.82	0.000
They	14 (0.23%)	658 (1.23%)	-71.35	0.000
Our	4 (0.06%)	426 (0.79%)	-65.86	0.000
Were	5 (0.08%)	422 (0.79%)	-60.48	0.000
Was	37 (0.60%)	714 (1.33%)	-28.95	0.000
Women	3 (0.05%)	209 (0.39%)	-27.70	0.000
Them	8 (0.13%)	297 (0.55%)	-27.08	0.000

Table 2 further distinguishes the two literary works by providing more supportive evidence of the number features, singular forms of pronouns and verbs. ‘I, is, me, am, it, my’ are keywords of positive correlation (287.19~28.36) while plural forms of ‘we, they, our, were, women, them’ are those of negative correlation (-124.82~-27.08) in the TYWP with the only exception ‘was’ (-28.95) included in the latter list. Since the differences reach the significant level, the result indicates that these words function greatly differently in the sense of usage frequency in the two works. ‘I’ is a high frequent word in Herland, as shown in Table 1, but it appears significantly less frequent (1.4%), compared to the word, ‘I’ in TYWP (4.94%). It means that there exists a strong contrast between the two texts with respect to the usage frequency of ‘I’, which plays a much more important role in TYWP than in Herland. Based on the distinction of pronouns observed between two works, we propose the Number Feature Hypothesis to differentiate them: elements of singular form in TYWP in contrast with those of plural form in Herland. Following the proposed hypothesis, we expect to find verbs in correspondence to the agreement of subject pronoun, i.e., plural forms of verbs should appear more frequently in Herland. However, there is one verb of singular form shown in the list, ‘was’ which belongs to keyword list of negative correlation for TYWP. That is to say, it appears abnormally high (1.33%) in Herland in contrast with the percentage of appearance (0.6%) in TYWP. The word ‘was’ frequently appears in Herland, as what has already been shown in Table 1. In order to test the Number Feature Hypothesis and clarify whether there are many subjects of singular form in Herland, we search the collocated elements of ‘was’, the result is shown in Table 3.

Table 3. Elements Collocated with ‘Was’

Order		TYWP		Herland
1	I	28	It	168
2	It	15	That	128
3	That	12	I	85
4	He	6	This	69

From Table 3, we discover the most frequent appearance of the collocated subject ‘it’ (number 2 in order) in TYWP compared to ‘I’ (number 3 in order) in Herland. The word, ‘I’, is the element with highest frequency collocated with ‘was’ in TYWP while ‘it’ and ‘that’ surpass ‘I’ in the text of Herland. The result accounts for the reasonable existence of ‘exceptional’-like for the case of singular verbal form ‘was’ which might become a counterexample for the proposed Number Feature Hypothesis at first glance. Hence, our classification hypothesis regarding number feature in the two texts stands after clarifying those subjects collocated with the singular ‘was’ in Herland.

Going one step further, we also derive another two keyword lists by comparing the two works with BNC separately, as shown in Table 4, which is different from Table 2 where they are first compared with the BNC separately while they are compared with each other in Table 2.

Table 4. Gilman's Two Texts Compared to BNC

P=0.000, from higher to lower positive correlation with BNC	
TYWP	I, me, am, my
Herland	We, our, us, women, they, motherhood, them, were, I, mothers, girls, men, theirs, ours, children, their, was, babies, child, mother, parentage, me, feminine, fatherhood, sisterhood, ourselves, woman, father, boys, man, motherliness, ladies, masculine

From Table 4, we can observe that singular forms found in keyword list of TYWP in contrast with BNC consist of 'I, me, am, my', while plural forms are found in the list of positive correlation of Herland compared to BNC, for example, 'we, our, us, they, them, were,...'. Based on the distinction and contrast shown by number feature (singular and plural) in the two texts, we associate the plural form with the concept of group/community, in contrast with the singular form, which is related to the individual. Hence, plural forms that entail the notion of 'group' are used to define the concept of unity and team in Herland, while singular forms in TYWP represent the individual concept.

The above analysis and discussion can be connected with existing literary study results about these two selected books. Firstly, the contrast of number features reflects that these two works are different genres: an autobiographical story which focuses on personal experiences and a utopian fiction where the greatest benefit to mankind is considered in this ideal world. Besides, number features indicate Gilman's presentation of women's questions in the two texts. TYWP shows a woman's predicament and the loneliness, solitude and worry of the principal protagonist are expressed while Herland exhibits a utopia ruled and managed by a collective group of women. Observing the singular pronoun in TYWP, we can notice that Gilman's intention to get a woman reveal the predicament of being a woman in her own voice. The repeated use of the first person pronoun reinforces the sympathy that female readers might have upon this female protagonist. Here Gilman points out the problem that every individual woman might confront as she tries to fulfill the role of a mother by constantly letting the narrator refer to herself. As for Herland, plural pronouns are frequently used because Gilman hopes to picture a world in which women unite and cooperate without the oppression of patriarchy. Unlike TYWP with which Gilman raises reader's awareness of women's predicament in patriarchal society, Herland further provides us with a possibility that women can work together to build an ideal society. In addition to shifting her focus from 'the individual' in TYWP to 'the collective' in Herland, Gilman discusses new themes in Herland than in TYWP.

4.2. The Gender Feature

Besides the number feature, we compare keywords in Gilman's texts and the balanced corpus, BNC. The result is shown as Table 4. By contrasting the high frequency words with those of a balanced corpus (BNC), Table 4 not only confirms the core distinction of number feature between the two literary works as mentioned in the previous section, but also indicates that gender issue is the principal theme in Herland but not in TYWP. To be more specific, lexicon associated with family issue particularly draws our attention. Related words including 'women, motherhood, mothers, girls, men, children, babies, child, mother, parentage, feminine, fatherhood, sisterhood, woman, father, boys, man, motherliness, ladies, and masculine' are listed as the keywords in the contrastive results of Herland and the huge corpus BNC. However, no related words are found in the keyword list of BNC and TWYP. The keyword function of WordSmith helps us to extract the main theme of Herland and distinguish Herland from TYWP in an efficient way. Based on above word lists and keywords, and further extension of expression patterns, we will discuss related themes in this section.

4.2.1. The Language Pattern

Following the previous discussion, we firstly focus on pronouns to search for repeated clusters in order to further observe language patterns. Some results are shown in Table 5.

Table 5. Clusters of Pronouns

Cluster	TYWP	(6,171 tri-gram)	Herland	(53,644 tri-gram)
Pronoun	I		we	
	I don't want	9 (0.0014%)	of course we	15 (0.0003%)
	of course I	5 (0.0008%)	we did not	11 (0.0002%)
	but I must	5 (0.0008%)	we are not	8 (0.0001%)
Pronoun+			we do not	8 (0.0001%)
			we had to	8 (0.0001%)
			we could not	8 (0.0001%)
			that we should	7 (0.0001%)
			we should have	6 (0.0001%)
			we were not	6 (0.0001%)
			why should we	5 (0.0001%)
Pronoun			they	
Pronoun+			they were not	20 (0.0004%)
			they had no	12 (0.0002%)
			they did not	12 (0.0002%)
			they are not	8 (0.0001%)
			they could not	5 (0.0001%)
			they are all	5 (0.0001%)
			of course they	5 (0.0001%)

From the searched clusters, it is interesting to discover the cluster of pronoun collocated with interpretation of resistance or insistence such as prepositional phrase ('of course'), modal verbs ('must, have to') and negation ('no, not'). Since clusters with negations have drawn our attention in the previous table, we further compare negation usages in TYWP and Harland. The result in Table 6 shows that negations appear similarly (1.6~1.8%) in the two texts studies. We will illustrate the relation between the expression patterns used in two texts and the intention implied by the author through examples.

Table 6. Comparison of 'Negation'

Negations	TYWP (6,171)	Herland (53,644)
Not	44	396
n't	53	195
No	14	260
	111 (1.80%)	851 (1.59%)

In TYWP, expressions of 'don't' show us that the protagonist is forbidden to do certain things, for example, to walk around ("Don't go walking" (p. 23)). Under her husband's supervision, the protagonist is afraid to reveal her real emotion. "I cry at nothing, and cry most of the time. Of course I don't when John is here, or anybody else, but when I am alone," says the narrator (p. 19). Through the following negative expressions, we see that she is trapped in a place which she dislikes as she states that "I don't like our room a bit" (p. 12), "I don't like it [the pattern] a bit. I wonder-I begin to think-I wish John would take me away from here!" (p. 22) and "You don't get me out in the road" (p. 35). Besides, the expressions of 'must (not)' in TYWP suggest that there are certain obligations for the protagonist which she 'must' or 'must not' do; that is, there is external coercion or authority that forces the protagonist to act accordingly. For example, she has the responsibility to control herself and gets well according to her husband, who says that "no one but myself [the narrator] can help me out of it, that I must use my

will and self-control and not let any silly fancies run away with me” (p. 22) and “But I must not think about that” (p. 16). Since she is forbidden to write, she “must put this [diary] away” because her husband “hates to have me [the narrator] write a word” (p. 13). In TYWP, the restriction that the society poses on women may be seen in men’s forbidding the narrator to do certain actions as well as the narrator’s negative reflection to the restriction delivered in the form of ‘don’t’ and ‘do not’. The clusters serve as examples to prove that TYWP aims to depict women as being denied, negated and patronized and to show how the demanding patriarchy poses the negative influence on the body and mind of women.

Negation can also be used to show some unexpected facts in terms of social behaviors. In Herland, the negative clusters, unlike in TYWP, show that women in Herland prove themselves to be different from how the male visitors thought them to be. That is, women in this female utopia don’t comply with the norms and images in a patriarch-centered society. For example, according to the narrator, these women “were not timid, inexperienced, weak” (p. 141), and “had no horrible ideas” (p. 111). Jens Allwood (1977) indicates that a negative statement tends to appear while its affirmative complementary is somehow anticipated (p. 8). These men in Herland say that Herlanders are ‘not weak’ first instead of ‘strong’ (p. 132). Therefore, we can infer that these men have the presupposition or stereotype that women are mostly weak while the characters of women in Herland are enforced by men through negation.

In order to observe more about the concept with respect to the expectation of women in Herland, derivations of words such as ‘expect’ and ‘suppose’ have been searched and related examples are shown and discussed below.

The three male visitors then compare women in their home society and those in Herland. While women in their home society were expected to have characteristics such as ‘pettiness’, ‘jealousy’, and ‘hysteria’, women in Herland have “a fair-minded intelligence” (p. 81); while women in their home society were expected to yield to ‘feminine vanity’ or ‘frills and furbelows’, women in Herland “evolved a costume more perfect than the Chinese dress, richly beautiful when so desired, always useful, of unflinching dignity and good taste” (p. 81); while women in their home society were expected to be “by nature conservative,” those in Herland were found different (p. 111). These are all stereotypes that traditional patriarchal society presupposes women to be. As for obligation, wives are expected by men to serve them and be loyal (p. 95) and all the women to be “dull submissive monotony” (p. 81). Men expect women to do the housework, considering it as inborn nature and pleasure. The narrator says in Herland that “the housekeeping duties and pleasures we, by instinct and long education, supposed to be inherently appropriate to women” (p. 123). Women are supposed to stay at home and men should be the ones that participate and compete in the world. Terry, one of the visitors, thinks that mothers “would of course work for their children in the home; but the world’s work was different—that had to be done by men, and required the competitive element” (p. 60).

Besides, women are expected to act appropriately and femininely in the 19th century. With respect to ‘femininity’, a virtue and norm for women set up by patriarchal society, which causes some traditional expectations, words such as ‘feminine’ and ‘masculine’ have been searched and results are discussed below. The word ‘feminine’ appears 15 times and ‘masculine’ 8 times in Herland. First, the traditional view of femininity is mentioned by Vandyck, one of the three male visitors to Herland and the narrator of the story. In the narration, he says that “You see, with us, women are kept as different as possible and as feminine as possible” and that “We men have our own world, with only men in it; we get tired of our ultra-maleness and turn gladly to the ultra-femaleness” (p. 129). Traditionally, women are supposed to be the opposite of men so that men can escape their maleness when they want to and from a male perspective, traditional femininity is depicted as being something ‘kept’ by men. For example, Jeff and Terry, two of the visitors in Herland, consider that women will be more feminine if they have long hair or do the housework like knitting (pp. 30-31). Not only does Gilman point out the existence of traditional femininity, she also hopes to explore how traditional femininity is constructed.

The traditional idea of femininity appears to her only as a contrast of masculinity. In the story the narrator says that “those ‘feminine charms’ we [men] are so fond of are not feminine at all, but mere reflected masculinity...” (p. 59).

In addition to femininity, Gilman also brings out new ideas related to masculinity. She shows us the traditional concept of masculinity as well as suggests us that masculinity can be unnecessary and useless. In *Herland*, traditional masculinity is depicted from a male perspective by three male visitors as being the center of the world. For instance, it is said in the text that “when we say men, man, manly, manhood, and all the other masculine derivatives, we have in the background of our minds a huge vague crowded picture of the world and all its activities” (p. 137). Gilman builds up a utopia *Herland* where the women have been living for a long time without the assistance of men. Furthermore, they do not have traditional femininity. The narrator is surprised at “how these ultra-women, inheriting only from women, had eliminated not only certain masculine characteristics, which of course we did not look for, but so much of what we had always thought essentially feminine” (p. 57). Gilman indeed provides us a new possibility besides masculinity and femininity.

In summary, various clusters and negation associate with the different intention, obligation in *TYWP* and expectation in *Herland*. It shows that women are deprived of certain freedoms and rights in *TYWP* and that in *Herland* these female *Herlanders* do not behave as what the patriarchal society has expected; they even reverse the hierarchy. From the results, we observe that Gilman’s focus has shifted from representing women’s predicament of being restricted by expectations to providing a possibility that women can escape all traditional expectations.

4.2.2. Mother

In order to analyze the theme of mother, we extract words derived from the lemma ‘mother’ for further observation because the lemma connotes the essential meaning of the word and the derivational forms contain the same base share the similar semantic feature. These related words include the plural form of ‘mother’, ‘mothers’, and nouns such ‘motherhood’, ‘motherliness’, and ‘mothering.’ We expect these derivational words to indicate elements associated with the studied themes since they convey the similar meaning of the base form of the word. Table 7 provides the comparison of usage frequency in three corpora.

Table 7. Distribution of Derivations of ‘Mother’

Sources Word	Herland Frequency (%)	TYWP Frequency (%)	BNC Frequency (%)
Mother	57 (0.1100)	1 (0.0200)	21,492 (0.0200)
Mothers	40 (0.0738)	X	3,225 (0.0000)
Motherhood	43 (0.0794)	X	296 (0.0000)
Motherliness	3 (0.0055)	X	2 (0.0000)
Mothering	1 (0.0019)	X	94 (0.0000)

Table 7 shows that derivations of ‘mother’ such as ‘mother’, ‘mothers’, ‘motherhood’, ‘motherliness’, and ‘mothering’ are words repeated and emphasized as a prominent lexical notion of the work *Herland* compared to two other corpora, *TYWP* and *BNC*. In *Herland*, the singular form ‘mother’ and the plural form of the word, ‘mothers’, appears 57 times (0.11%) and 40 times (0.0738%), the word, ‘motherhood’, 43 times (0.0794%), ‘motherliness’ 3 times (0.0055%), and ‘mothering’, once (0.0019%). The other related words are ‘mother-feeling’, ‘mother-love’, and etc. Most of them are repetitively used in *Herland*, while in *TYWP*, except the word, ‘mother’, appearing once, there is no other word related to the meaning of mother. The result indicates the fact that the word ‘mother’ has a much higher frequency in *Herland* (0.11%) than that in a natural corpus, *BNC* (0.02%). The extraordinary high proportion of a word appearing in single literary work signifies the author’s intention to draw readers’

attention and to impress them as well.

Besides the quantitative description, we are going to analyze the meanings of words with lemma 'mother' (including 'mothers', 'motherhood' and 'motherliness') by studying related phrases and their contexts. All examples have been analyzed except for the one used as interjection "Mother of Mike, boys -- what Gorgeous Girls! To climb like that! to run like that! and afraid of nothing. This country suits me all right. Let's get ahead." (Gilman, 1979, p. 17). Even the mother-related terms show the identical tendency of Gilman's use of contrast to the two stories and the mother figures, the biological and the social, we are going to examine them in order to verify our assumption that Gilman's effort to challenge the traditional concept of motherhood. There are eight categories derived: (1) mother as a person who gives birth; (2) mother as gender, a female or a woman; (3) mother as a familial and social identity, a parent to a child; (4) mother as a profession, to care and to educate a child; (5) mother as a person with inborn or instinct ability to take care of a child; (6) mother as an actual character in the novel; (7) mother as a person with a loving and caring trait, (8) mother as a religious character. The distributional inclination of meanings contained in the related words is shown in Table 8.

Table 8. Meaning Classification of Derivations of 'Mother' in Herland

Category	'Mother'	'Mothers'	'Motherhood'
(1) a person who gives birth	18	12	22
(2) gender, a female or a woman	5	2	12
(3) a familial and social identity, a parent to a child	11	18	1
(4) a profession, to care and to educate a child	3	5	0
(5) a person with inborn or instinct ability to take care of a child	3	0	0
(6) an actual character in the novel	10	1	0
(7) mother as a person with a loving and caring trait	4	0	0
(8) a religious character	2	2	0
(1) a person who gives birth & (2) gender, a female or a woman	0	0	8

Table 8 demonstrates that the word, 'mother', except once used as an interjection, actually refers to different ways to view mother as a role or identity. In category (1), we can see the word 'mother' is used to refer a person who has given birth. The other categories with higher frequency are category (3) and (6). The result listed in Table 8 shows that the plural form of the word 'mother' is used similar as its singular form. That is, both the biological and the social functions of a mother are emphasized by frequent usage. While both functions are vital for Herlanders, they somehow respectively designate these two duties to different groups of Herlanders. Some have the ability to give birth, while some others nurture the babies collectively. Furthermore, 'motherhood' is mostly used to refer as the power of 'virgin birth', which is the most vital element for the survival of this women community. Thus, in Table 8, as we can see that 'motherhood' is used to imply a woman's ability to give birth more frequently than to indicate a social duty as a caretaker of a child that a woman is often assigned to, there is a discrepancy.

The result also echoes with previous literary studies which indicate that motherhood is one of the themes concerned in Herland. We can infer that in Herland the emphasis is not only on the biological function of a woman, but also on her relationship as a parent to a child at home and in a society. Traditional motherhood gives no clear distinctions between what a mother should do to fulfill her role. Therefore, this kind of motherhood inflicts all the responsibility of a woman, which makes motherhood a horrifying task. In Herland, the idea of mother has been divided into several sectors, each of which is essential and interrelated. Due to this clear categorization, Herlanders demonstrate a new motherhood that is achieved through cooperation and teamwork. The above analysis shows that Gilman emphasizes not only the mother but also the new motherhood in a quantitative and semantic way.

The various meanings generated from the word, mother, not only indicate the importance of mother Gilman wants to emphasize, but also bring about a very interesting contrast to *TYWP* where the main protagonist, an impotent mother who is troubled with depression, is trapped in her own psychological predicament. The protagonist in *TYWP* is not capable of undertaking the task of being a mother who nurses and educates a child, though she is still the biological mother. Therefore, the role of 'mother' disappears in the work. The term 'mother' is only used once and is not even referred to the female protagonist in the story, but to her mother-in-law instead. Both works are related to the issue of motherhood and a mother's social and familial roles, but the ways Gilman demonstrates her thinking by deploying the key term 'mother' are completely opposite. Here, we conclude that by purposely avoiding the word 'mother' in *TYWP*, Gilman points out that traditional motherhood is pathological and thus cannot be called as a mother. By analyzing meanings of derivations of 'mother' in *Herland*, the responsibility of mother is shared by women in the *Herland* while the principal protagonist in *TYWP* suffers by herself, we are able to confirm with Gough (1995)'s central statement, mother is a multifaceted and complex role, which should not be simplified and generalized.

We have demonstrated linguistic features and patterns related to number and gender features found in the two works through Tables 1-8 to answer our research question 1 (What are the linguistic features and patterns found in the two works?). Furthermore, with respect to research question 2 (How is the language use in the two works related to the themes of the works?), we have illustrated the relation between the language use in the two works and their themes such as concepts of group/individual and feminine/family with examples and interpretations based on the data analyzed. Finally, the quantitative and qualitative analysis of language usage helps better illuminate previous discussions of various themes such as 'expectation' and 'mother' by literary critics. Through corpus-based approach, not only words and patterns of high frequency, but also non-prominent features can be detected, for example, 'mother' in *TYWP*.

5. Conclusion

This study has identified the linguistic features and patterns of Gilman's two works, *TYWP* and *Herland* through corpus tools. The linguistic features (number and gender) have been related with two themes, "the group and the individual" and "the feminine and the masculine". Firstly, the obvious difference between the two works has been analyzed in this study in contrast with the reference corpus, pronouns are high frequent words in those two works while this is not the case for the general usage in a balance corpus. Secondly, the contrast of number indicates that first-person singular forms used more frequently in *TYWP* to portrays women as isolated individuals who must fight a lone battle and plural ones in *Herland* to depict women as a united group whose self-confidence overturns patriarchy. Thirdly, words related to gender such as 'women, motherhood, mothers, girls, men, feminine, fatherhood, woman' in *Herland* show that gender issue is the principal theme in *Herland* but not in *TYWP*. Besides, the collocation of pronoun and negation, suggesting the obligatory command of patriarchal society is observed in *TYWP* from the narrator's use of language, while the negation in *Herland* reveals a subversion of traditional values regarding women. Lastly, with proof of linguistic analysis of 'mother' related words, we have illustrated Gilman's sophisticated perception of the role of mother. In conclusion, a corpus-based approach to literary study facilitates a more efficient literary analysis and provides further evidence for textual analysis in literary texts. Future research could include discourse analysis to enhance our findings, and apply the corpus-based approach to other literary works by Gilman in order to reach a more comprehensive conclusion.

Acknowledgments

This work was supported by the Landmark Project of National Cheng Kung University under grant number D100-3100.

References

- Allwood, J. (1977). Negation and the strength of presuppositions: Or there is more to speaking than words. In Dahl, O. (Ed.), *Logic, Pragmatics and Grammar* (pp. 11-52). Lund: Student Literature.
- Dock, J. B. (Ed.). (1988). *Charlotte Perkins Gilman's "The Yellow Wall-paper" and the History of Its Publication and Reception: A Critical Edition and Documentary Casebook*. University Park, PA: Pennsylvania State University Press.
- Fischer-Starcke, B. (2010). *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*. New York, NY: Continuum.
- Gilman, C. P. (1892). *The Yellow Wall-paper and Other Writings*. New York, NY: Bantam Dell.
- Gilman, C. P. (1979). *Herland*. New York, NY: Pantheon Books.
- Gilman, C. P. (1935). *The Living of Charlotte Perkins Gilman: An Autobiography*. New York, NY: D. Appleton-Century Company.
- Gilman, C. P. (1981). *The Yellow Wallpaper*. London: Virago.
- Gough, V. (1995). Lesbians and Virgins: The New Motherhood in Herland. In Seed, D. (Ed.), *Anticipations: Essays on Early Science Fiction and Its Precursors* (pp. 195-206). Liverpool: Liverpool University Press.
- Harbour, D., Adger, D., & Béjar, S. (2008). *Phi Theory: Phi-features across Modules and Interfaces*. New York, NY: Oxford University Press.
- Hardy, D. E. (2007). *The Body in Flannery O'Connor's Fiction: Computational Technique and Linguistic Voice*. Columbia: University of South Carolina Press.
- Hedges, E. R. (1973). Afterword. In Gilman, C. P., *The Yellow Wallpaper* (pp. 37-73). London: Virago Press.
- Kerstens, J. (1993). *The Syntax of Number, Person and Gender: A Theory of Phi-features*. Berlin: Mouton De Gruyter.
- Lane, Ann J. (1990). *To Herland and Beyond: The Life and Work of Charlotte Perkins Gilman*. New York, NY: Pantheon Books.
- Leech, G. (1992). Corpora and theories of linguistic performance. In Svartvik, J. (Ed.), *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter.
- Mahlberg, M. (2010). Corpus linguistics and the study of nineteenth-century fiction. *Journal of Victorian Culture*, 15(2), 292-298.
- Ross, R. (1973). The coefficient of concordance: A mathematical tool for comparing literary texts. *Style* 7(1), 1-20.
- Schörkhuber, V. (2008). *Charlotte Perkins Gilman's The Yellow Wallpaper: An Analysis*. Norderstedt: Druck und Bindung.
- Sveen, H. A. (2005). "Honourable" or "Highly-sexed": Adjectival Descriptions of Male and Female Characters in Victorian and Contemporary Children's Fiction. Stockholm: Elanders Gotab.
- Weatherford, J. (1999). Approaching the ineffable: "The yellow wallpaper" and Gilman's problem with language. *American Studies in Scandinavia*, 31, 58-75.

THE AUTHORS

Hui-Chuan Lu is currently a full professor at Department of Foreign Languages & Literature, National Cheng Kung University, with emphasis in Spanish linguistics, corpus linguistics and syntax studies.

Kai-Ling Liu is a researcher the Department of Foreign Languages & Literature, National Cheng Kung University, with emphasis in 19th century American Literature and women's studies.

Chien-Ting Yeh completed Master's degree at the Department of Foreign Languages & Literature and was a member of Multilingual Corpora Team at National Cheng Kung University.

Ya-Jie Chen completed Master's degree at the Department of Foreign Languages & Literature and was a member of Multilingual Corpora Team at National Cheng Kung University.

THE AUTHORS' ADDRESSES**First and Corresponding Author****Hui-Chuan Lu**

Professor

Foreign Languages and Literature

National Cheng Kung University

No.1, University Road, Tainan City, 70101, TAIWAN

E-mail: huichuanlu1@gmail.com

Co-author**Kai-Ling Liu**

Researcher

Foreign Languages and Literature

National Cheng Kung University

No.1, University Road, Tainan City, 70101, TAIWAN

Email: kailing@mail.ncku.edu.tw

Chien-Ting Yeh

MA Student

Foreign Languages and Literature

National Cheng Kung University

No.1, University Road, Tainan City, 70101, TAIWAN

Email: ctyeah1020@gmail.com

Ya-Jie Chen

MA Student

Foreign Languages and Literature

National Cheng Kung University

No.1, University Road, Tainan City, 70101, TAIWAN

Email: yajiec@hotmail.com

Received: 15 April 2022

Received in Revised Form: 10 August 2022

Accepted: 21 August 2022