A Corpus Stylistic Study of Singular and Plural Keywords in Jane Austen's *Persuasion*

RAKSANGOB WIJITSOPON

Department of English, Faculty of Arts, Chulalongkorn University, Bangkok 10330, Thailand
Raksangob.W@chula.ac.th

Abstract. While many corpus stylistic studies that draw on keywords pay attention to meanings of the lexical items, the present study focuses on their grammatical forms and argues that the forms largely account for the significance of keywords of a text and their contributions to meanings in a literary text. The nouns "years" and "man", which were found to be statistically significant in Jane Austen's *Persuasion*, were studied in terms of their co-occurrence patterns and textual functions. Concordance lines of these two words were analysed in comparison with their singular and plural non-key counterparts. The analysis reveals that the singular and plural forms of the keywords play an important role in the construction of thematic meanings and narrative techniques in the novel. It is suggested that an examination of grammatical aspects of keywords can provide contributions not only to literary studies of a novel but also to corpus stylistic methods and to theoretical linguistic arguments on word meanings.

Keywords and phrases: corpus stylistics, keyword, lexis-grammar interface, *Persuasion*, singular-plural forms

Introduction

Over the past decade, there have been an increasing number of corpus stylistic studies that investigate the ways in which language is used to create meanings and effects in literary texts through reliance on a corpus, i.e., a large collection of electronic texts sampled to be maximally representative of a particular language or language variety (McEnery and Hardie 2012). This is not only because the corpus linguistic approach makes it possible to examine a whole long text such as a novel or even a group of literary works but also because it can shed light on some linguistic features and patterns that are unlikely to be found through careful reading or even through manual textual analysis.

One of the central descriptive tools in corpus stylistic research that can lead to the discovery of hidden linguistic patterns that account for literary interpretations is the concept of the "keyword". The term "keyword" refers to lexical items that have significance in a given text because they are indicators of the text's content or style or because of their "unusual frequency in comparison with a reference corpus of some suitable kind" (Scott and Tribble 2006, 55). In other words, a word form that is often repeated within the text under investigation is very likely to be indicative of the text's meaning and/or style. Identifying keywords often serves as a starting point for an analysis because it points to textual elements that are potentially meaningful in the creation, interpretation and organisation of the text.

In fact, the definition of keyword in corpus linguistics cited above overlaps with the concept of "style markers" in stylistics, which are defined as "linguistic features whose densities in a text are appreciably different from those found in its contextually related norm" (Enkvist 1973, 21). The references to "frequency" (keyword) and "densities" (style marker) suggest that quantification plays an important part in the identification of both style markers and keywords. This in turn points to an underlying assumption that there is a correlation between lexical repetition and perceptual and statistical significance (Scott and Tribble 2006, 58). The other overlapping characteristic of the concepts keyword and style marker is that comparison between a text and a particular type of norm is central to identifying keywords and style markers. This is reflected in Enkvist's (1973) and Scott and Tribble's (2006) references to a "contextually related norm" and "a reference corpus of a suitable kind", respectively. The comparative method is premised upon the rationale that a text is a selection from the potential of a linguistic repertoire, and hence, a comparison between a text and a contextually related norm of language use represented by a corpus allows us to study how far a given text consists of recurrent linguistic patterns that are widespread in the language as a whole or that represent a relevant language variety (Stubbs 2005, 21).

While there have been quite a number of corpus stylistic studies that draw on keywords, almost all of them pay attention to the semantics of keywords and the way this relates to different aspects of the literary texts under study, for example, thematic meanings and fictional world (e.g., Mahlberg and McIntyre 2011), characterisation (e.g., Culpeper 2014a, 2014b; Jaafar 2017) and cohesion (e.g., Mastropierro and Mahlberg 2017). To illustrate this practice with corpus stylistic work on Jane Austen novels, Starcke (2009), for example, approaches keywords in Jane Austen's *Pride and Prejudice* by categorising and grouping them in light of their semantic relationships with one another, such as mental concepts and emotion, women and family relationships. These semantically related groups

of keywords were then analysed to determine in what ways they account for certain interpretations concerning family relationships in the novel.

Although it cannot be denied that words and their semantic properties are closely related, it has been noted through a great deal of corpus linguistic research, such as Römer and Schulze (2009), that lexis and grammar are inseparable and that together they are responsible for the construction of meaning in communication. The grammatical properties of a keyword, however, have rarely been addressed in corpus stylistic analysis. Unlike many previous studies, the present study, therefore, takes a different approach to keywords, focusing on their grammatical forms, specifically on the singular and plural forms, as contributing linguistic factors in some keywords' interpretative and functional relevance to Jane Austen's final novel *Persuasion*. This also entails a comparison between some keywords and their non-key counterparts, which have similar meanings but different forms. Through the comparison, the ways in which grammatical forms of keywords are related to their keyness status and interpretations of the novel can be clearly seen.

To this end, I first give a synopsis of and background information about *Persuasion*, followed by a theoretical explanation of a corpus linguistic perspective on the lexis-grammar interface. Then, a description of methodological procedures that lead to a list of keywords in the novel is provided before two keywords are chosen for detailed analysis, i.e., "years" and "man". These two keywords are analysed in comparison with their non-key singular and plural counterparts to examine in what ways their grammatical forms are associated with the keyness of the lexical items and with the textual meanings of the novel. Finally, I give concluding remarks on the ways in which an investigation into the grammatical properties of keywords can benefit stylistic analysis and discuss potential contributions of the findings to future work in corpus stylistics as well as to theoretical linguistic descriptions of word meanings.

The Novel *Persuasion*

Persuasion is Jane Austen's last novel, published posthumously in 1817. The novel tells the story of a 27-year-old woman named Anne Elliot, who rejected a marriage proposal made by the man she loves, Frederick Wentworth, when she was 19 years old, at the persuasion of her late mother's dear friend who felt that Wentworth "has nothing but himself to recommend him" (Austen 1817, 41). Since then, Anne has been heartbroken and lived with grief and feelings of loss. Eight years later, she and Wentworth, who is now a successful and wealthy naval captain, happen to be in the same circle of friends and acquaintances. Captain Wentworth is determined to marry a woman with "a strong mind" (Austen 1817, 68), whose

decisions and opinions cannot be easily swayed by others. However, he gradually comes to appreciate Anne's value through a number of incidents before they finally reveal the love for each other that still remains.

Because *Persuasion* is Jane Austen's last novel, it has often been discussed in literary studies in comparison with other novels by the same author. While Austen's works are generally considered comedies (cf. e.g., Page 1972; Kirkham 1997), *Persuasion* has been observed to be "the most […] sober in color of all Jane Austen's novels" (Jenkyns 2004, 196). This is probably because the novel deals mainly with the heroine's endurance, suppressed grief and loneliness and because the novel frequently references mishaps or accidents involving injuries and death. All these together create a perilous fictional world unlike any found in Jane Austen's other novels (Todd 2015). The main protagonists are also often felt to be different from Jane Austen's other heroes and heroines. Anne is considered to be fully equipped with birth, intelligence and wealth and yet is "invisible" for the first part of the novel, suffering from the greatest emotional impoverishment (Jenkyns 2004, 2). At the same time, the novel presents us with a hero interested in money, who boasts about his progress in life and bears a grudge against the heroine.

In addition, unlike other Austen novels, *Persuasion* has been observed to contain references to domestic and international political situations of the time, for example, to the Napoleonic War. Discussion of the novel also surrounds socio-cultural characteristics of the period. Social mobility is an example. The fact that the hero is a self-made man, having acquired prestige and wealth by joining the navy, is a reflection of the changing social mobility in Regency England (McMaster 1997). Gender issues also feature in literary discussions of the novel, such as a redistribution of gender characteristics (Jordan 2007) and a feminism that highlights women's strength (Todd 2015). Jordan (2007, XIV) even argues that *The Strengths of Women* could have been an alternative title for *Persuasion*.

As reflected by these literary remarks, *Persuasion* seems to have a number of features that distinguish it from its predecessors. It is the focus of this study to explore how its distinctive qualities are manifested linguistically and textually and what patterns of language use contribute to its distinctive style when compared with other novels by the same author.

Lexis-Grammar Interface: A Corpus Linguistic Perspective

Research in corpus linguistics has consistently emphasised the lexis-grammar interface, i.e., how lexis and grammar are intertwined in the construction and

interpretation of meaning in language. Different senses of a word are usually distinguished by the structures in which the word occurs, whereas different words that have a similar grammatical pattern tend to have a semantic property in common (cf. e.g., Sinclair 1991; Hunston and Francis 2000). What leads to these observations are findings from various corpus studies that bring out one of the significant phenomena in language that does not feature a distinction between lexis and grammar: "co-selection", i.e., words habitually occurring in repeated forms of the verbal environment (cf. Sinclair 1991). In other words, meanings and uses of lexical items are closely tied to their patterns of co-occurrence. The co-textual features define the context of the occurrence of a core expression at the lexical, grammatical, semantic and pragmatic levels. These four dimensions of co-occurring patterns can be translated into the following four descriptive categories:

1. Collocation: In this descriptive model, collocation is used broadly to refer to a concrete component of co-textual features. The concept draws on the repetition of words on the textual surface around the node word/phrase. In other words, collocation is viewed as a text-based concept, defined explicitly in Sinclair (1991, 20): "collocation is the occurrence of two or more words within a short space of each other in a text". By this definition, collocation is a consequence of the linearity of language. Based on this definition, collocates of a node word/phrase can be extracted manually from an examination of concordance lines.

2. Colligation: According to Sinclair (2004, 32), colligation refers to a pattern of co-occurrence between a pair of grammatical instances or between a node word and grammatical categories. For example, over 90 percent of the phrase "naked eye" in Sinclair's analysis co-occurs with these words: "with", "to", "by" and "from", each of which can be referred to as a collocation, or a collocate, of "naked eye". However, as can be seen, they all share the same part of speech, i.e., preposition. Therefore, the word class "preposition" is a colligation of the phrase "naked eye". Colligation is thus seen as a step further in abstraction than collocation: grammatical relations between the node word and a set of collocates are assigned after the collocates of a node word/phrase are identified.

3. Semantic preference: This is a relation between a node word/phrase and a set of semantically related words. For example, Sinclair (2004) observes that the words "see" and "visible" mostly co-occur with the phrase "naked eye". Other words that occur less frequently in the same environment include: "detect", "spotted", "perceived", "apparent", "evident" and "obvious". These collocates share, roughly, a common semantic field

of "visibility". Hence, it can be said that the phrase "naked eye" has a semantic preference for "visibility". In this respect, semantic preference is relatively closely tied to the concept of collocation in that it describes a phenomenon whereby a particular lexical item collocates frequently not with another item but with a series of items that belong to a particular semantic set.

4. Semantic prosody: Semantic prosody expresses the pragmatic dimension of a lexical item. It plays a crucial role in the creation of meaning in a text, as it helps suggest how a lexical item is to be interpreted functionally in its context. As Sinclair (2004, 34) states, "without it, the string of words just 'means' – it is not put to use in viable communication". The concept, however, has been defined differently in various studies (cf. e.g., Stewart 2010; Whitsitt 2005). The present study draws upon the conceptualisation given in Louw's study (1993). According to Louw[1] (1993), whose article introduced the concept to the public for the first time, semantic prosody is an evaluative or attitudinal tone a word articulates. This colouring capacity, however, is not a built-in property of a word in the way that connotation is but results from the habitual co-occurrence between the word and a set of words that share similar semantic traits (Partington 2004). Hence, semantic prosody may not even be accessible to an individual's introspection or recorded in dictionaries but may only be evident through concordancing numerous examples. For example, in Louw's (1993) analysis, the word "utterly" is considered as having a negative semantic prosody because it frequently collocates with words with negative meanings.

As seen from the principle of co-selection and types of co-occurrences above, grammar is part of a network of co-occurrence patterns between lexical items, realised explicitly in the category "colligation" and contributing to units of meaning in a text. It should be noted here that the four descriptive categories above will be applied as an analytical framework in the keyword analysis below.

The form-meaning relationship is manifested not only through the co-occurrence patterns of a lexical item but also through different grammatical forms of a lemma or lexeme[2], represented by capital letters in the present study. For instance, it has been demonstrated that different grammatical forms of the noun lemma EYE contribute to semantic variations of the word (Sinclair 2003); the plural form "eyes" tends to express the literal meaning of the word as part of the body, as seen through its occurrence with colour terms such as "blue eyes" or "bruised eyes"; on the other hand, the singular form "eye" tends to be used figuratively or as part of an idiom,

as often realised through such phrases as "sharp eye" or "the naked eye". Stubbs (2002) suggests that this form-meaning relationship of a word form can be tied to text and discourse. This is illustrated through his analysis of different forms of the verb lemma SEEK, i.e., "seek", "seeks", "seeking" and "sought", which shows that while these forms are often taken as grammatical variations, they are used in very different ways. The word "seeks" has no shared collocates with any of the other forms, except with "seek", and their only shared collocate is "professional". This suggests that "seeks" and "seek", generally regarded as having the same meaning, actually tend to be used in different ways and express different sorts of units of meaning. Additionally, while "seek", "seeking" and "sought" are often found in political and legal contexts, "seeks" is found in advertisements and news headlines. Grammatical forms are thus intertwined with lexical items, meanings and properties of texts. Stubbs' (2002) observation, however, is based on data from a general corpus, which comprises a variety of text types, as it addresses issues in lexical semantics. In what ways grammatical forms of a word are relevant to properties of an individual text, for example, its meanings and organisation, is of interest in this paper.

It is these theoretical concepts that underlie the present stylistic analysis of some keywords in *Persuasion*. An examination of the distinction between the keywords selected for a thorough investigation and their singular/plural counterparts will be conducted to determine how the grammatical form of each keyword helps in an articulation of textual meanings of *Persuasion*.

Methodology and Preliminary Results

To discover what lexical items mark the style of *Persuasion* and how patterns of these words account for the novel's meanings, an electronic version of *Persuasion* was downloaded from the Project Gutenberg website (https://www.gutenberg.org) and compared with a corpus of Jane Austen's five other major novels (henceforth JA5), which contains *Northanger Abbey*, *Sense and Sensibility*, *Pride and Prejudice*, *Mansfield Park* and *Emma*. This is because JA5 is the most "contextually related" possible norm for a comparison with *Persuasion* to extract style markers of the novel, as the novels were all written by the same author. Furthermore, as mentioned above, discussions of *Persuasion* often involve a comparison with its predecessors, so using JA5 as a parameter for the comparative purpose in stylistics can yield linguistic insight into an aspect of literary interest concerning differences between the author's last novel and her earlier ones.

It might be of use to note that, apart from a corpus of Jane Austen's other novels, a corpus of 19th century British novels, which contains none of Jane Austen's

novels, might be seen as an alternative reference corpus for a stylistic analysis of *Persuasion*. However, upon a pilot comparison between *Persuasion* and a corpus of 19th century British novels, it was found that keywords derived from the comparison were mostly lexical items reflecting Jane Austen's authorial style rather than the textual style of *Persuasion* itself; the keywords that turned up were in almost exactly the same set as those found in Wijitsopon (2013), who investigated the style of Jane Austen's six major novels, for example, "be", "very", "she" and "must". Therefore, only JA5 was used as a reference corpus, as it is the closest and most contextually related reference corpus for a comparison with *Persuasion*. *Persuasion* and JA5 contain 83,367 and 693,861 tokens, respectively.

The software used for the extraction of keywords in the present study is Wmatrix (Rayson 2003), which provides standard corpus linguistic methodologies such as frequency lists and concordances. It also offers a web interface to the English semantic and part-of-speech corpus annotation tools. This allows for an extension of the keyword method to that of key grammatical categories and key semantic domains.

To extract keywords in *Persuasion*, criteria were set as follows:

1. A minimum frequency of 30 was chosen for a lexical item to appear on the keyword list. This choice came as a consequence of multiple experiments on an optimal minimum frequency for keyword extraction in the present study. It was found that items that occur less than 30 times tend to be names of minor characters or places that are related to the novel at a specific point in the plot, e.g., "Winthorp", "Shepherd", "tenants", "nurse" and "concert", rather than being used throughout the novel.

2. The statistical measure log-likelihood (LL), a default calculation measure on Wmatrix, was adopted. The LL value of 10.83, which amounts to the p-value of ≤ 0.001, was set as a cut-off point. This is because after several experiments, it was found that a higher cut-off point would yield only items specific to the fictional world of *Persuasion*, e.g., major characters' names and titles as well as places, while lower ones were found to allow too many lexical items to become "key".

Based on these criteria, a total of 18 lexical items that are not proper nouns or titles (e.g., Sir, Admiral) turn up as keywords in *Persuasion*. These keywords are shown in Table 1.

Table 1. Keywords in *Persuasion*, excluding proper nouns and titles

No.	Keyword	Frequency	Relative frequency	LL value
1	had	1165	1.45	138.93
2	been	495	0.61	42.7
3	the	3217	3.99	29.71
4	they	434	0.54	28.84
5	and	2764	3.43	25.58
6	he	961	1.19	22.94
7	years	47	0.06	22.7
8	were	426	0.53	21.17
9	all	462	0.57	21.05
10	found	81	0.1	16.94
11	being	214	0.27	16.33
12	acquaintance	61	0.08	14.84
13	seemed	90	0.11	14.19
14	man	134	0.17	13.24
15	among	39	0.05	12.51
16	near	40	0.05	11.89
17	those	72	0.09	11.37
18	others	58	0.07	11.26

Of the 18 keywords above, 13 are function words, and 5 are content words. The fact that more than twice as many function keywords as content keywords turn up is probably because function words are essential nuts and bolts in the creation and organisation of texts. However, as the present study is interested in the form–meaning correlation of keywords in a text, more attention was paid to content keywords. This in no way suggests that function words are not important. As Groom (2010, 59) argues, function words can "form a valid and even preferable basis for empirical linguistic research into specialised discourses". A future stylistic analysis of this novel can pursue the ways in which the function keywords in Table 1 above interact with the text of *Persuasion* in the production of meanings and effects. After all, it should be noted that some top-ranking function keywords in the list above, such as "had" and "been", have been explored. For example, Starcke (2006) looks at the patterns of a most frequent phrase "she had been" in *Persuasion* and links it to the temporal structure of the novel, while Wijitsopon (2013) addresses variants of the verb lemma BE in Austen's novels, including

"been" and "be", and their different collocational patterns in an expression of the novels' thematic concerns.

The five content keywords in *Persuasion* are "years", "found", "acquaintance", "seemed" and "man". Given that these are the top five key content words, after proper nouns and titles, and that they can "compete" with high-frequency function words and appear in the keyword list, their occurrences can be deemed highly significant to meanings of the novel, and hence, some of them are pursued further in the present study. The noun "years" is the most key of all. Importantly, it should be noted that only the plural form is key, while its singular counterpart does not turn up on a keyword list, even when the cut-off point is set lower. This leads to the hypothesis that its significance to the text under study is attributed not only to its meaning but also to its being plural. This word is therefore chosen for further analysis in the section below. The other keyword to be investigated is "man" because, like "years", it is considered a general noun (Mahlberg 2005), which is normally used for a referential purpose in discourse. It would therefore be interesting to examine its usage patterns as to how a general noun such as "man", with its singular form, contributes to meanings of *Persuasion* when the plural form "men" does not turn up on the keyword list. The other content keywords are not examined in the present study for the following reasons. Only the singular form "acquaintance" occurs in the novel, whereas its plural form "acquaintances" does not occur at all. Hence, there are no data for examining how different forms of the lemma ACQUAINTANCE are used in the novel. The verbs "found" and "seemed" were not chosen for analysis because the past and past participle forms of these keyword verbs are far more associated with the fictional type of the text under study than with the contribution to meanings specific to this novel. In other words, as fictional stories are generally narrated in the past tense, the forms of the above keyword verbs point to the characteristic of *Persuasion* as fictional prose, rather than to the textual meanings and potential effects of the novel. Hence, they seem less relevant to the purpose of the present study than the singular and plural forms of the noun keywords.

Analysis

To illustrate how grammatical forms of the keywords in question account for textual meanings of *Persuasion*, both the singular and plural forms of the noun lemmas MAN and YEAR were analysed. This is so that it can be seen very clearly why and how the singular or plural form accounts for each individual word's keyness and for its relations to the novel.

YEAR: "year" – "years"

The frequency of the two forms of YEAR is as follows: "year" occurs 27 times in the novel and "years" 47 times. Of the 27 instances of the singular noun "year", 18 cases (66.67%) are embedded in prepositional phrases, altogether denoting the duration of an event/action or when something happens. For example:

> - Captain Harville had taken his present house <u>for half a year</u>.

> - That occurred before I went to sea <u>in the year six</u>,"

The other eight instances function as an adverbial phrase, for example, "She had been at Bath the <u>year</u> before". There is only one case of "year" that occurs in the subject position: "The <u>year</u> of their engagement could not but be named by him".

Based on this formal analysis, it can be said that the singular noun "year" mainly serves to signal the temporal orientation of events and actions in the novel. The sample concordance lines below indicate that events and actions encoded in collocation with the time noun "year" are often those related to "staying or living somewhere", "going somewhere", "waiting or expecting" and "marrying or engagement", typical actions found in Austen's novels (Wijitsopon 2013).

of and <u>expected</u> all the rest of the	**year**	; but he never came . The following
nt living , <u>found a home</u> for half a	**year**	, at Monkford . He was, at that tim
; and it being now too late in the	**year**	<u>for such visits</u> to be made on foot ,
at the time ; but I believe about a	**year**	<u>before he married</u> Mary . I wish she
ea . <u>They were come</u> too late in the	**year**	for any amusement or variety which
could not but be reverted to ; the	**year**	<u>of their engagement</u> could not but be

Figure 1. Concordance lines of "year" in *Persuasion*

On the other hand, the occurrences of "years" in *Persuasion* not only serve to express time orientation regarding a period of more than one year in the narrative text but also are specifically linked to textual meanings related to sadness. This is illustrated by the fact that, while 25 out of 47 instances (53.19%) of the keyword serve to indicate the time duration of events that occur, in the same way as its singular counterpart does, a total of 22 of the 47 instances of "years" (46.81%) co-occur with words and expressions related to tragic events, misfortune, loss and decline, as exemplified in Figure 2.

been followed by so many, many	**years**	of <u>division and estrangement</u>. There
as still, after an interval of several	**years**,	<u>felt with anger</u> by Elizabeth, who had
every one sooner than myself. Six	**years**	of <u>separation and suffering</u> might have
er, but she felt her approach to the	**years**	<u>of danger</u>, and would have rejoiced to
think of her as he would. No: the	**years**	which <u>had destroyed her youth and bloom</u>
hat the other had imagined. Twelve	**years**	<u>had changed Anne from the blooming</u>, sil
were invariably gentle; and twelve	**years**	<u>had transformed the fine-looking, well-</u>
kwardness and its emotion. Twelve	**years**	were gone <u>since they had parted</u>, and
avagant; and <u>at his death</u>, about two	**years**	before, had left his <u>affairs dreadfully</u> invol

Figure 2. Concordance lines of "years" in *Persuasion*

Based on these patterns, the plural noun "years" is found to perform two functions in the text, i.e., to indicate time duration and to provide a contextual temporal link to various kinds of ordeals in the novel. In other words, the plural form "years" is also used to suggest a long period of time in which characters suffer, as suggested by its collocation patterns above, and hence can be said to express negative semantic prosody of a period of more than 12 months. The keyness of "years" is thereby associated with its plural form in that it contributes to an expression of thematic notions concerning time, memory, endurance and sadness in the novel (Todd 2015), in a way that the singular form does not.

Apart from collocation and semantic prosody, an investigation into the colligational patterns of "years" also reveals why the plural noun features significantly in *Persuasion*. In looking at all 47 instances of "years", it was found that the noun is preceded by different types of determiners, as shown in Table 2.

Table 2. Colligational patterns of "years" in *Persuasion*

Type of pre-determiner	Frequency (47)	Example
Numeral	27 (57.45%)	The man who understands somewhat of the value of an Anne Elliot, should, <u>eight years</u> afterwards, be charmed by a Louisa Musgrove.
Quantifier	17 (36.17%)	The offence which had been given her father <u>many years</u> back, she knew; […]
Definite article	2 (4.26%)	But she felt her approach to <u>the years</u> of danger.
No determiners	1 (2.13%)	Captain Wentworth, repeating his name so often, puzzling <u>over past years</u>, […]

As shown in Table 2, the most dominant colligation of "years" in *Persuasion* is numerals. While it is possible to assume that a reference to "years" in a literary text does not necessarily indicate an actual length but a prolonged period of time, the dominance of numerals as the major colligation of "years", which give precise information about the length of time, seems particularly interesting. It can be interpreted that specific references to the length of "years" constitute an important stylistic feature of the novel.

A close look at the textual positions in which these 27 instances of the "numeral + years" colligational pattern occur reveals that a number of them occur repeatedly in the presentation of characters' thoughts about the past. The references to the length of years reflect that they have not forgotten and are still affected by past events. This can be illustrated through an excerpt below, which presents Anne's thoughts after encountering Captain Wentworth for the first time since they parted.

> Soon, however, she [Anne Elliot] began to reason with herself, and try to be feeling less. Eight years, almost eight years had passed, since all had been given up. How absurd to be resuming the agitation which such an interval had banished into distance and indistinctness! What might not eight years do? Events of every description, changes, alienations, removals--all, all must be comprised in it, and oblivion of the past--how natural, how certain too! It included nearly a third part of her own life.
>
> Alas! With all her reasoning, she found, that to retentive feelings eight years may be little more than nothing.
>
>   (Austen 1817, 67; underline by author)

In the extract above, the repetition of the phrase "eight years" contributes to the narrator's free indirect presentation of Anne's thoughts and, at the same time, it reflects her obsession with her past relationship. While it can be argued that this colligational pattern is dominant due to such repetitions, the reiteration of the phrase is stylistically significant in that it reflects the protagonist's preoccupation with her failed relationship in the past and at the same time registers her emotional agitation upon meeting Captain Wentworth again.

Apart from being used in the presentation of a character's mental and emotional state, the colligation pattern "numeral + years" is also often used in the narrator's satirical description of some characters. References to the number of "years" function as an intensifier of time duration in which something happens. In the extract below, the phrases "ten years" and "thirteen years" are mentioned to show

that even though a long time has passed, Sir Walter Elliot is still vain about his eldest daughter's appearance.

> It sometimes happens that a woman is handsomer at twenty-nine than she was ten years before; and, generally speaking, if there has been neither ill health nor anxiety, it is a time of life at which scarcely any charm is lost. It was so with Elizabeth, still the same handsome Miss Elliot that she had begun to be thirteen years ago, and Sir Walter might be excused, therefore, in forgetting her age, or, at least, be deemed only half a fool, for thinking himself and Elizabeth as blooming as ever, amidst the wreck of the good looks of everybody else; for he could plainly see how old all the rest of his family and acquaintance were growing. Anne haggard, Mary coarse, every face in the neighbourhood worsting, and the rapid increase of the crow's foot about Lady Russell's temples had long been a distress to him.

> (Austen 1817, 25; underline by author)

The satire above originates from the fact that Sir Walter Elliot still believes that he and his eldest daughter, Elizabeth, look as glamorous now as they did 13 years earlier. The precise reference to the number of years and such hyperbolic expressions as "as blooming as ever, amidst the wreck of the good looks of everybody else" pragmatically create a series of exaggerations, reflecting Sir Elliot's vanity and at the same time hinting at the narrator's satirical depiction of the character.

Another case of the way in which the colligation "numeral + years" serves to exaggerate a long period of time for a satirical purpose is found in the narrator's description of Elizabeth's position as the mistress of Kellynch Hall.

> Elizabeth did not quite equal her father in personal contentment. Thirteen years had seen her mistress of Kellynch Hall, presiding and directing with a self-possession and decision which could never have given the idea of her being younger than she was. For thirteen years had she been doing the honours, and laying down the domestic law at home, and leading the way to the chaise and four, and walking immediately after Lady Russell out of all the drawing-rooms and dining-rooms in the country. Thirteen winters' revolving frosts had seen her opening every ball of credit which a scanty neighbourhood afforded, and thirteen springs shewn their blossoms, as she travelled up to London with her father, for a few weeks' annual enjoyment of the great world.

> (Austen 1817, 26; underline by author)

The stylistic parallelism created through the repetition of "thirteen years", in tandem with the reiteration of "thirteen" in collocation with "winters" and "springs", both of which also denote periods of time, in the thematic position of the clauses (Halliday and Matthiessen 2004) foregrounds the length of time in which Elizabeth has enjoyed a luxurious life as the mistress of Kellynch Hall, which in turn helps to explain her dismay and disappointment with her situation later in life.

To summarise, while the singular form of "year" is mostly used to indicate the time in which something happens in *Persuasion*, the collocational and colligational patterns of the plural noun "years" in the novel suggest that it not only denotes a period of time longer than one year but also creates subtle textual meanings related to the novel's melancholic tone and the characterisation of major characters as prisoners of time. Furthermore, it serves as a lexical tool for a stylistic presentation of a character's thoughts and the creation of sarcastic remarks in the fictional world of *Persuasion*. This corresponds to Sinclair's (2003) argument that, while it is often thought that the only difference in meaning between the singular–plural forms is that of grammatical number, there are cases wherein different grammatical forms can constitute different units of meaning. In the case of the present study, it is argued that the plural form of the lemma YEAR serves functionally to create textual meanings and stylistic effects in *Persuasion*.

MAN: "man" – "men"

As shown in Table 1, the singular form "man", with its frequency of 134 tokens and the LL value of 13.24, is on the list of the top 18 keywords, while its plural form is not. A comparative examination of the concordance lines for both words reveals that although the two forms of the lemma MAN refer to male characters, they have different collocational patterns in *Persuasion* and thereby reflect different contributions to the novel.

Drawing upon Mahlberg's (2005, 104) framework, which focuses on the discourse function of general nouns denoting people, an investigation into the concordance lines for the singular and plural forms of MAN in *Persuasion* reveals that the nouns "man" and "men" are used in two cases:

1. Referring to a specific character or group(s) of characters:

    - Sir Walter Elliot, of Kellynch Hall, in Somersetshire, was a <u>man</u> who, for his own amusement, never took up any book but the Baronetage;

    - One day last spring, in town, I was in company with two <u>men</u>, striking instances of what I am talking of;

2. Referring to a type, not an individual character or group of characters:

  - We shall never agree, I suppose, upon this point. No <u>man</u> and <u>woman</u>, would probably.

  - If there is anything disagreeable going on, <u>men</u> are always sure to get out of it.

Based on these functional categories, the distribution of the two grammatical forms and functions of MAN can be spelled out in Table 3.

Table 3. Distribution of referential functions of "man" and "men" in *Persuasion*

Man (134)		Men (31)	
Referring to a specific character	Referring to a type, not an individual	Referring to specific characters	Referring to a type, not individual characters
93 (69.40%)	41 (30.60%)	10 (32.26%)	21 (67.74%)

It can be seen from Table 3 that the singular form tends to be used to refer to a particular character more than a generic kind, while the generic use applies more to the plural form. This is probably because, as observed by Mahlberg (2005), the plural form tends to denote a more general meaning of words than does its singular counterpart. A further analysis of concordance lines reveals why specific and generic uses of the keyword "man" are dominant in the novel.

Of 93 instances of the specific use of "man", 70 cases (75.27%) co-occur with evaluative adjectives or phases, as illustrated in the underlined parts of the concordances below:

fifty-four , was still <u>a very fine</u>	**man**.	Few women could think more of their
lter -- " " He would be <u>a very lucky</u>	**man**,	Shepherd , " replied Sir Walter , "
<u>a very hale , hearty , well-looking</u>	**man**,	a little weather-beaten , to be sur
that time , <u>a remarkably fine young</u>	**man**,	with a great deal of intelligence ,
Elliot is <u>an exceedingly agreeable</u>	**man**,	and in many respects I think highly
ore . He is <u>a clever man , a reading</u>	**man**	-- and I confess that I do consider h
are unprejudiced . Mr. Elliot is a	**man**	<u>without heart or conscience</u> ; a desig
uld discern Mr. Smith to have been a	**man**	<u>of warm feelings , easy temper , care</u>

Figure 3. Concordance lines of the specific use of "man" in *Persuasion*

This pattern reflects that references to specific male characters are often made in the novel because male characters are consistently assessed in terms of their personal attributes. Although it has been observed that characters in Austen's novels make constant judgements of one another (e.g., McMaster 1995), this most dominant pattern of the keyword "man" in *Persuasion* suggests that in this novel, making judgements on an individual male character is featured more than it is in her other novels. On the one hand, this may be because the novel contains a variety of male characters, including sailors of different ranks, aristocrats and noblemen. On the other hand, this pattern can suggest that, while Jane Austen's novels are generally seen as concerning women, *Persuasion*, on the basis of its being compared with JA5, highlights the story of men more than the other Austen novels; their character and lives are important to the plot and progression of the novel. Furthermore, while literary discussions tend to pay attention to feminism and the theme of female power in *Persuasion*, as noted earlier, the keyness of "man" and its collocational patterns suggest a point that perhaps deserves more attention in literary discussions of the novel, i.e., Austen's treatment of male characters in *Persuasion*.

The keyness of "man" is also attributable to the generic use of "man" and its co-occurrence patterns. Although its generic function is found less frequently than the specific type, it cannot be denied that 30.60% of the occurrences of "man" constitute the keyness of the singular form. A total of 18 out of 41 instances of the generic singular "man" (43.90%) display a striking collocational pattern in that the noun tends to co-occur with the word "woman" or words related to relationships with women, for example, "love", "attachment" or "marrying". In other words, the generic use of "man" in *Persuasion* has a semantic preference for "woman" and for the domain related to "man-woman relationships". This can be illustrated in the concordance lines below:

for half a year together? If a	**man**	had not a <u>wife</u>, he soon wants
ed in aid here. In <u>marrying</u> a	**man**	indifferent to me, all risk would any
When one lives in the world, a	**man**	or <u>woman's marrying</u> for money
ure . I will not allow it to be more	**man**	's nature than <u>woman 's</u> to be inconst
sert that ours are the most tender .	**Man**	is more robust than <u>woman</u> , but he is
s and a half ago . Dare not say that	**man**	forgets sooner than <u>woman</u> , that his as

Figure 4. Concordance lines of the generic use of "man" in *Persuasion*

These co-occurrence patterns of the keyword "man" can be seen as adding textual evidence, in the form of collocational patterns, to literary discussions of gender issues in the novel, in addition to a reference to a particular scene at a particular point in the novel, often cited by literary critics. Given that *Persuasion* is compared with other major Austen novels in the present study, the keyness of the singular noun "man" leads us to see that the author's final work foregrounds the presentation of male characters' qualities and man-woman relationships more explicitly than her previous novels. In other words, the keyness of "man" is thus accounted for by an amalgam of specific and generic uses of the word, which plays an important role in the characterisation and expression of textual meanings regarding man-woman relationships.

Let us now take a look at the plural form of MAN, which, as mentioned above, displays a clearly different co-occurrence pattern from that of "man". The majority of the generic "men", i.e., 16 out of 21 (76.19%), collocate with words related to comparison, especially "other", as illustrated in the concordance lines below:

rse , and few navy officers , or	**men**	of any other description , can have h
equal claim with any other set of	**men**,	for all the comforts and all the pr
utmost : I know no other set of	**men**	but what lose something of their pers
ght on the subject more than most	**men**.	" <div1 type=chapter n=8> From th
and warmth than any other set of	**men**	in England ; that they only knew ho
t was better to look at than most	**men**,	and he had no objection to being se
ld not divide her more from other	**men,**	than their final separation. Pretty reless
immethodical , like other	**men**,	about those things ; and when I cam
and just rewards . Like other great	**men**	under reverses , " he added with a sm

Figure 5. Concordance lines of the generic use of "men" in *Persuasion*

Common collocational patterns of "men" in this novel thus include "other set of men", "most men" and "other men". This pattern suggests that the generic plural noun "men" occurs mainly in a comparative discourse, wherein male characters in *Persuasion* are divided into different groups and evaluated comparatively. A closer look at the textual environment larger than the concordance lines reveals that the comparison is often made between naval officers and other groups of men, such as aristocrats. This collocational pattern of the generic use of "men" in *Persuasion* can be seen as a set of linguistic evidence that illustrates the novel's treatment of

the issue of the rise of self-made men. Critics' observations on social mobility during that period are thereby rendered not only by the characterisation of Captain Wentworth and the inclusion of a number of sailor characters in the fictional world but also by linguistic patterns in the novel. Although the concept of career is a notable point in the novel, as often remarked on by literary critics, for example, McMaster (1997), the keyness of "man" over "men" and the disparity in their collocational patterns suggest that evaluation of male characters and man-woman relationships dominate the fictional world of *Persuasion*.

A point worthy of note concerning the generic use of "man" and "men" in *Persuasion* is that, even though both forms were found to be used in a generic sense, they have different patterns of semantic preference. The former tends to collocate with the semantic domain related to relationships with a woman, whereas the latter tends to collocate with the semantic domain regarding career. The grammatical distinction between the two forms of MAN in this novel is thus also largely accompanied with a distinction in semantic domains that constitute the text of *Persuasion*.

Conclusion

The present study approaches keywords in a literary text in a way that is different from other corpus stylistic studies. It highlights a grammatical, rather than a semantic, property of the given keywords. This involves a comparison between key and non-key lexical items with similar meanings but different grammatical forms. Through this new approach to keywords, it has been shown that the grammatical forms of lexical items are closely related to the status of keywords and to their contributions to meanings and effects of the literary text in which they occur. As has been demonstrated, the plural form of the keyword "years" in *Persuasion* is linked to meanings and semantic prosody of the noun in that it suggests a long time of suffering, which contributes to the novel's theme of endurance. The singular form of "man", though not obviously affecting meanings of the lexical item as was seen in the case of "years", is used for purposes different from "men". That is, it is part of the referential function in major narrative techniques adopted in the novel, i.e., the author's characterisation of male characters and her explicit thematic instantiation about man-woman relationships throughout *Persuasion*.

While findings from the present study suggest the importance of a grammatical perspective in keyword analysis, it is illustrated through an analysis of only two pairs of nouns. Future corpus stylistic work can shed more light on a grammatical category as a key linguistic feature of a literary text, instead of treating it as a property of keywords in the way the present study has demonstrated. This can be done by extending the concept of "keyword" to "key grammatical category" by

grammatically annotating a literary text using such software as Wmatrix (Rayson 2003), as mentioned above.

In addition, while the present study views a classic literary work from the linguistic perspective and proposes an alternative approach to keyword analysis in corpus stylistic research, it must also be noted that the findings from this study can lead us to see the interplay among lexis, grammar and text. The different semantic prosodies associated with "year" and "years" and the semantic preferences for different domains that "man" and "men" have are related to the fictional world of the novel under study. It can thus be said that lexis and grammar interact not only with each other in the construction and interpretation of meanings but also with the textual dimension of language use. This is in line with Hanks' (2000, 214) argument on word meanings:

> Words have meaning potentials, rather than just meaning. The meaning potential of each word is made up of a number of components, which may be activated cognitively by other words in the context in which it is used. […] So rather than asking questions about disambiguation and sense discrimination ("Which sense does this word have in this text?"), a better sort of question would be "What is the unique contribution of this word to the meaning of this text?"

The present study has shown that the uses and "meaning potentials" of the nouns under study are subject to their grammatical forms and co-textual features in the text of *Persuasion*. Further analyses of other texts or text types will provide more empirical evidence for the integration of lexis, grammar and text in our approach to meaning in language. A corpus stylistic analysis thus not only yields linguistic evidence for literary interpretations but also can provide a basis and evidence for the development of theoretical linguistic descriptions.

Acknowledgements

Notes

1. The concept "semantic prosody" first appeared in Louw (1993). However, he states that the term and the concept were first used by Professor John Sinclair in their personal communication.

2. According to Stubbs (2002, 25), the word "word" is ambiguous since it may refer to an individual lexical item with a different inflectional form, e.g., "does", or to a set of morphological variants like DO, which includes different word forms, "do", "does", "did", "doing" and "done".

References

Austen, J. 1817. *Persuasion*. London: Pan Books Ltd.

Culpeper, J. 2014a. Keywords and characterization: An analysis of six characters in *Romeo and Juliet*. In *Digital literary studies: Corpus approaches to poetry, prose and drama*, eds. D. Hoover, J. Culpeper and K. O'Halloran, 9–34. London: Routledge.

_____. 2014b. Developing keyness and characterization: Annotation. In *Digital literary studies: Corpus approaches to poetry, prose and drama*, eds. D. Hoover, J. Culpeper and K. O'Halloran, 35–63. London: Routledge.

Enkvist, N. 1973. *Linguistic stylistics*. The Hague: Mouton. https://doi.org/10.1515/9783111348926

Groom, N. 2010. Closed-class keywords and corpus-driven discourse analysis. In *Keyness in Texts*, eds. M. Bondi and M. Scott, 59–78. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.41.05gro

Halliday, M.A.K. and Matthiessen, C. 2004. *An introduction to functional grammar*. Oxon: Hodder Education.

Hanks, P. 2000. Do word meanings exist? *Computers and the Humanities* 34: 205–15. https://doi.org/10.1023/A:1002471322828

Hunston, S. and Francis, G. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.4

Jaafar, E. 2017. Corpus stylistic analysis of Thomas Harris' *The Silence of the Lambs*. *Khazar Journal of Humanities and Social Sciences* 20(1): 25–42.

Jenkyns, R. 2004. *A fine brush on ivory: An appreciation of Jane Austen*. Oxford: Oxford University Press.

Jordan, E. 2007. Introduction. In *Persuasion*, J. Austen, VII–XV. London: Wordsworth Classics. https://doi.org/10.1177/0957155807078028

Kirkham, M. 1997. *Jane Austen, feminism and fiction*. Exeter: Short Run Press.

Louw, W. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In *Text and technology: In honour of John Sinclair*, eds. M. Baker, G. Francis and E. Tognini-Bonelli, 157–176. Amsterdam: John Benjamins. https://doi.org/10.1075/z.64.11lou

Mahlberg, M. 2005. *English general nouns: A corpus theoretical approach*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.20

Mahlberg, M. and McIntyre, D. 2011. A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction* 4(2): 204–227.

Mastropierro, L. and Mahlberg, M. 2017. Key words and translated cohesion in Lovecraft's *At the Mountains of Madness* and one of its Italian translations. *English Text Construction* 10(1): 78–105. https://doi.org/10.1075/etc.10.1.05mas

McEnery, A. and Hardie, A. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McMaster, J. 1995. *Jane Austen the novelist*. London: Palgrave McMillan.

_____. 1997. Class. In *The Cambridge companion to Jane Austen*, eds. E. Copeland and J. McMaster, 115–130. Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521495172.007

Page, N. 1972. *The language of Jane Austen*. Bristol: Basil Blackwell.

Partington, A. 2004. *Patterns and meanings*. Amsterdam: John Benjamins.

Rayson, P. 2003. Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. PhD dissertation, Lancaster University, United Kingdom.

Römer, U. and Schulze, R. 2009. *Exploring the lexis–grammar interface*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.35

Scott, M. and Tribble, C. 2006. *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.22

Sinclair, J. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

_____. 2003. *Reading concordances*. London: Longman.

_____. 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.

Starcke, B. 2006. The phraseology of Jane Austen's *Persuasion*: Phraseological units as carriers of meaning. *ICAME Journal* 30: 87–104.

_____. 2009. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics* 14(4): 492–523. https://doi.org/10.1075/ijcl.14.4.03fis

Stubbs, M. 2002. *Words and phrases*. Oxford: Blackwell.

_____. 2005. Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature* 14(1): 5–24. https://doi.org/10.1177/0963947005048873

Stewart, D. 2010. *Semantic prosody: A critical evaluation*. Oxon: Routledge.

Todd, J. 2015. *Cambridge introduction to Jane Austen*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316178591

Whitsitt, S. 2005. A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics* 10(3): 283–305. https://doi.org/10.1075/ijcl.10.3.01whi

Wijitsopon, R. 2013. A corpus-based study of the style in Jane Austen's novels. *Manusya: Journal of Humanities* 16(1): 41–64.