The quality of choices determines the quantity of Key words

Punjaporn Pojanapunya and Angvarrah Lieungnapar
King Mongkut's University of Technology Thonburi

Abstract

Keyword analysis is a corpus-based technique which identifies important words, or keywords, in a given corpus as compared against a benchmark corpus. To identify words that are considered as key for interpretation (Key words), researchers are required to make choices including the number of texts in a target corpus, a benchmark corpus, and a cut-off of top words in a keyword list. Different choices will produce different keyword lists. The problem is that it is unclear to what extent the choices will influence the keyword lists. This study intends to highlight the effects of these choices by conducting several simulations of a target corpus of the discussion section of research articles. The simulations vary in the number of texts in the target corpus, type of benchmark corpora, and number of top words. A property of Key words was identified and compared across different settings. The results show that the benchmark corpus influences both number and type of Key words, while the number of texts in the target corpus affects only word ranking, and the cut-off of top words affects the number of Key words upon condition. Consideration of these methodological choices is discussed.

1. Identifying Key words

Keyword analysis has been used as a stepping-stone for several kinds of analyses, especially corpus-based analysis of texts. It has been used in various disciplines, both in and outside applied linguistics, e.g. in English for Specific Purposes (Kwary, 2011), register and stylistics (Freddi, 2005), project management (Crawford, Pollack, & England, 2006), and medical sociology (Seale, 2008). It is widely used because the method provides keyword lists which can indicate important concepts and styles of the texts or corpora under investigation (Scott, 2009).

In essence, to identify keyword lists, analysts start by building a corpus of texts for investigation, also known as a target corpus. Normally, keywords in keyword lists are defined as words which occur with a significantly high frequency in a target corpus when measured against a comparative corpus. The next step to identify keywords is to choose a comparative corpus, or a reference corpus, to be used as a benchmark for word frequency comparison between the two corpora (Rayson, 2013; Scott, 2010). By using specific statistics for keyword analysis, a number of keywords that tend to occur more frequently in a target corpus when compared with its relative frequency in a reference corpus (Scott, 1997; Scott, 2010; Scott & Tribble, 2006) can be automatically identified.

However, due to the large number of keywords provided by the analysis, not all of the identified keywords in keyword lists can be instantly used. Instead, to be able to focus on a manageable number of keywords in keyword lists, a cut-off point of top words needs to be set. The shortened keyword lists identified from the cut-off of top words provide analysts with top keywords that are manageable for investigation within individual studies. These words are referred to as "Key words" in the present study. Identification of these Key words allows analysts to meet the purposes of individual studies. However, identifying Key words can be problematic. Although keyword analysis is an automated method, it requires analysts' decision-making in almost all of the steps of analysis. For instance, analysts have to select the type of texts for investigation. Selection of text type can be relatively straightforward, depending on research purposes or research questions of individual studies. However, the

decisions on some other factors influencing Key words identification can be less straightforward.

There seem to be three main factors that must be taken into consideration when conducting keyword analysis: (1) How many texts should be collected to build a target corpus?; (2) What comparative or reference corpus should be used?; and (3) How many top words should be selected for a keyword list? Decisions on these three factors can be problematic, because several factors must be considered in making them. It is noted in the literature that different decisions on different choices can result in different keyword lists (e.g. Paquot & Bestgen, 2009; Scott, 2009). This suggests that choices in keyword analysis seem to influence the identification of Key words. However, to what extent these choices influence the keyword lists is still under-researched.

This study, therefore, intends to highlight how different choices influence Key words identification through the following factors in keyword analysis: (1) number of texts in a target corpus, (2) type of comparative corpus, and (3) number of top words in a keyword list as the cut-off. The results from this study will address the following two research questions:

1. Do choices of the number of the texts in a target corpus, type of benchmark corpora, and number of top words affect the proportion of Key words in a keyword list?
2. How do these choices affect the properties of Key words?

2. Methodology

2.1 The target corpus

The present study aims to investigate if choices on the numbers of the texts in a target corpus, types of benchmark corpora, and numbers of top words affect the proportion of Key words in a keyword list. To allow us to investigate all of these choices, we need a target corpus consisting of texts that meet the following criteria. First, to be able to show different results produced from target corpora varying in size or number of texts, we need texts which are similar in length and are available for large-scale retrieval. Second, to be able to compare the target corpus with different comparative corpora, a target corpus needs to have properties which can be compared against a comparative corpus varying in contextual scope. Therefore, the target corpus used in this study was a collection of discussion sections of research articles (D) in applied linguistics journals.

2.2 The Settings

To produce lists of Key words from several simulations, choices of three main factors that were investigated in this study are explained as follows:

1. Choices of the number of texts in a target corpus (no. of texts) being investigated are 20, 60, 100, and 400 texts. Based on a quick survey of 20 recent items of research investigating research article data, most of them investigated at least 20 texts (e.g. in Bruce, 2009; ElMalik & Nesi, 2008; Kuteeva & McGrath, 2013) with approximate numbers of texts ranging from 60 to 80 research articles (e.g. in Kanoksilapatham, 2007, 2012; Wannaruk & Amnuai, 2015). Therefore, 20 was set as the minimum number of texts, 60 as a standard quantity, 100 as an extended number of the average range, and 400 as a very large number of texts compared to the evidences from previous research.
2. Choices of the types of comparative corpus (benchmark) being tested are a content-different corpus, a corpus of full text, and a corpus of general English. These three types of comparative corpora can represent three types of comparisons: (1) the comparisons between a target corpus and a comparative corpus differing in content (a corpus of Introduction sections or I), (2) differing in scopes of texts in a collection (Introduction, Methodology and Results sections of the same research articles or IMR), and (3) differing in genre types (the British National Corpus or BNC representing general English).

3. Choices on the cut-off rankings (top n) being investigated are the top 25, 50, 100, and 200 keywords. When considering the number of top words in a keyword list, there are some general guidelines for researchers such as a keyword list should not be overwhelming in number of keywords for identification (Culpeper, 2009), the number should be adequate to classify keywords into themes, and the number should provide a variety of word types. Our preliminary survey revealed the three most common cut-off rankings of keywords which are top 25 (e.g. Seale, Ziebland, & Charteris-Black, 2006; Freddi, 2005), top 100 (e.g. Gerbig, 2010; Kang & Yu, 2011), and top 10 (e.g. Kotzé, 2010; Jimarkon & Watson Todd, 2013). The top 10 cut-off, however, seems to be too few to classify words into groups in the analysis of this study. Therefore, top 25 and top 100 were set as the testing cut-off rankings, with top 50 as the middle number. Additionally, top 200 was also used as the maximum number of keywords for the investigation.

Table 1 presents a summary of simulation types on choices in three factors varying in the number of texts (no. of texts), type of comparative corpora (benchmark), and cut-off rankings of words (top n).

Table 1. Descriptive information of simulation studies

No	Factors in keyword analysis	Choices	Types of studies	No. of texts	Benchmark	Top n
1	Number of texts in D (no. of texts)	*20, 60, 100, 400*	1a	-	(BNC)	(100)
			1b		(IMR)	(100)
			1c		(BNC)	(200)
2	Comparative corpus (benchmark)	*I, IMR, BNC*	2a	(100)	-	(100)
			2b	(100)		(50)
			2c	(60)		(100)
3	Cut-off rankings (top n)	*Top 25, 50, 100, 200*	3a	(100)	(BNC)	-
			3b	(100)	(IMR)	
			3c	(60)	(BNC)	

Although three factors were investigated in this study, each simulation type was set to investigate one factor with one condition at a time. Therefore, two default factors were determined for each of the simulation studies, namely benchmark and top n for study type 1, number of texts and top n for study type 2, and number of texts and benchmark for study type 3. The default settings for the three factors are: texts = 100, benchmark = BNC, Top n = 100. These settings are presented in brackets in Table 1.

To investigate the individual factors, Table 1 presents testing choices for the three factors in *italics*, and two types of studies (a main study and a checking study) were conducted. First, 1a, 2a, and 3a are the main studies conducted to investigate various numbers of texts, benchmarks, and top n, respectively. Furthermore, to check if the results of the main study were affected by the observed choice rather than chance, two other checking studies (b and c types of studies) for each of the simulation types were conducted.

For example, simulation 1 consists of three studies. The main study (1a) was conducted to investigate whether different numbers of texts (20, 60, 100, 400) would have an effect on the resulting Key words when the default factors use the BNC as benchmark and top 100 as the cut-off ranking to identify Key words. If the results of four Key word lists generated for the study when investigating 20, 60, 100, or 400 texts (1a) are different, it is possible that these results are affected by the use of the BNC or the cut-off rankings instead of the observed factor (no. of texts). Therefore, 1b and 1c studies were conducted as the checking studies. Compared with 1a, the 1b study differs in the default benchmark, while the 1c study differs in the default cut-off rankings. The results of these two checking studies were then compared with the main one to increase the opportunity that the conclusion would be drawn based on the influence of the observed factor rather than the other two.

2.3 Procedure and analysis

All simulations varying in number of the texts in the target corpus, type of benchmark corpora, and number of top words as shown in Table 1 were conducted. For comparing the Key words from different conditions, we need to describe them through some meaningful characteristics. Previous research using keyword analysis method has generally grouped, described, and discussed Key words through various aspects mainly based on specific criteria which serve particular research purposes. For example, Pojanapunya and Watson Todd (2016) classified Key words into words associated with research focus, research paradigms, and presentation and interpretation of results, to discuss identities of the discussion sections of research articles in soft and hard sciences. Carreon, Watson Todd, and Knox (2013) studied the growth of medical tourism through an analysis of private hospital websites. They classified Key words into words related to medical technology, cosmetic services, and business matters. However, as a broad level of Key word classification, literature in keyword analysis suggests that Key word results can generally be grouped into "aboutness" words (or words indicating what a corpus is about) and other closed-class words (Scott & Tribble, 2006), and they become the broad groups which are applicable to all keyword research. In this study, therefore, words in several keyword lists generated from different simulations were classified as (1) aboutness words and (2) other closed-class words by the two researchers to ensure inter-rater reliability. Next, proportions of aboutness Key words across different settings were compared and discussed.

3. Results and implications

To investigate choices on the three observed factors: number of texts, benchmark, and top n, Key words indicating aboutness (called as aboutness Key words in this study) were coded by the two researchers (rate of agreements is 94.9%); the number of aboutness Key words was counted and calculated into percentage for each of the studies. The following section shows proportions of the aboutness Key words of simulations for investigating numbers of texts (no. of texts) in Table 2, for investigating types of comparative corpora (benchmark) in Table 3, and for cut-off rankings (top n) in Table 4.

3.1 Study 1: Investigating numbers of texts in a target corpus (no. of texts)

The three studies as shown in Table 2 were conducted to examine whether or not the number of texts in a target corpus would have an effect on characteristics of Key words. First, focusing on the main study (1a), proportions of Key words indicating aboutness are generally similar (73.0-78.0%) although the number of texts in a target corpus varied between 20, 60, 100, and 400. When using IMR as a benchmark corpus in the 1b study, proportions of aboutness could also be considered similar (42.0-51.0%), although they are slightly lower than those of the main study. Since these results could also be affected by the third factor, we therefore

included more Key words for classification at 200 words in the 1c study. The results show that the aboutness Key words were fairly similar to the 1a study, with a range from 65.0 to 73.0 percent. As a whole, the findings are likely to show that the number of texts in the target corpus does not affect the Key word results.

Table 2. Proportions of the aboutness Key words in the studies varying with the number of texts in a target corpus

Studies		Lists	No. of texts	Benchmark	Top n	Aboutness (%)	General Interpretation
Main study	1a	1	*20*	(BNC)	(100)	73.0	No difference
		2	*60*	(BNC)	(100)	75.0	
		3	*100*	(BNC)	(100)	76.0	
		4	*400*	(BNC)	(100)	78.0	
Cross-check: benchmark	1b	1	*20*	(IMR)	(100)	51.0	No difference
		2	*60*	(IMR)	(100)	42.0	
		3	*100*	(IMR)	(100)	47.0	
		4	*400*	(IMR)	(100)	48.5	
Cross-check: Top n	1c	1	*20*	(BNC)	(200)	65.0	No difference
		2	*60*	(BNC)	(200)	71.5	
		3	*100*	(BNC)	(200)	72.0	
		4	*400*	(BNC)	(200)	73.0	

Top 100 Key words in all four lists as the results of 1a were observed. We found that many words overlapped across the four lists (e.g. *learners*, *language*, *knowledge*, *students*, and *English*); however, they differ only in their rankings. While *learners* is in the first rank in all the lists, *language* is ranked fifth in the lists using 20, 100, 400 texts in the target corpus, and ranked sixth in the list using 60 texts. *Knowledge* is ranked ninth, 18th, 16th, and 14th when we investigated 20, 60, 100, and 400 texts, respectively. Apart from words related to English Language Teaching (ELT) as given in the examples, it is also noticeable that many Key words shared among these four lists were research-oriented, e.g. *discussion*, *findings*, *results*, and *participants*.

The results of this simulation type suggest that the number of texts in the target corpus does not really affect proportions of the aboutness Key words, with approximately more than 70 percent in each list. With double the number of Key words identified (top 200 Key words), similar proportions of aboutness were again found in the list. Therefore, if researchers are interested in aboutness Key words, we suggest that it is possible to use a small number of texts where appropriate, with the minimum number at 20.

3.2 Study 2: Investigating types of comparative corpora (benchmark)

Study 2 was conducted to study the influence of types of comparative corpora used for keyword analysis on the proportion of aboutness Key words. The main study (2a) shows that using different types of benchmark seems to have a substantial effect on the number of aboutness Key words, e.g. 43.0 percent in a comparison between D and I, and 76.0 percent in a comparison between D and BNC. The investigations also show similar patterns of results in the 2b and 2c studies, where the top n was changed from 100 to 50 words, and number of texts was reduced from 100 to 60 texts in the target corpus, respectively. We can see that 2b condition contains 40.0 to 80.0 percent, whereas 2c contains 42.0 to 75.0 percent of aboutness Key words. In summary, types of comparative corpora influence proportions of aboutness keywords when the conditions differed both in terms of top n (2b) and number of texts (2c).

Table 3. Proportions of aboutness Key words in the studies varying by type of benchmark

Studies		Lists	Benchmark	No. of texts	Top n	Aboutness (%)	General interpretation
Main study	2a	1	*I*	(100)	(100)	43.0	Difference
		2	*IMR*	(100)	(100)	47.0	
		3	*BNC*	(100)	(100)	76.0	
Cross-check: top n	2b	1	*I*	(100)	(50)	40.0	Difference
		2	*IMR*	(100)	(50)	44.0	
		3	*BNC*	(100)	(50)	84.0	
Cross-check: no. of texts	2c	1	*I*	(60)	(100)	43.0	Difference
		2	*IMR*	(60)	(100)	42.0	
		3	*BNC*	(60)	(100)	75.0	

When considering Key words in the three lists of the 2a study, we found that comparing D with I (used as the corpora having the same contextual scope and size) produced several research terms (e.g. *discussion*, *results*, *conclusion*, *participants*, and *implications*). When comparing with the BNC which is used as the corpus different in terms of size and genres, the dominant words are related to ELT (e.g. *learning*, *language*, *students*, *proficiency*, *vocabulary*, and *comprehension*). This means that the use of different benchmarks not only affects the proportions of aboutness Key words, but also produces different types of Key words reflecting different aspects of the target corpus. Additionally, in terms of the proportion of aboutness Key words, the more similar benchmark to the target corpus (i.e. I) commonly produces fewer aboutness Key words than using the more different benchmark (i.e. BNC).

These results suggest that when researchers use benchmarks similar to a target corpus (e.g. D vs. I), they should consider identifying Key words in lower rankings (higher top n) to obtain a desirable proportion of aboutness Key words. On the other hand, when comparing a target corpus with a large difference between the two corpora (e.g. D vs. BNC), it is possible to look at less n or at fewer Key words on the top of the list.

3.3 Study 3: Investigating cut-off rankings (top n)

This last set of studies was conducted to investigate whether or not cut-off rankings would have an effect on the proportion of aboutness Key words. First, focusing on the main study (3a) classifying top 25, 50, 100, and 200 as aboutness Key words and other words, the results show that the list of top 25 Key words contains the most aboutness (96%), and the proportion decreases when more Key words are analyzed through the list. This means that closed-class words are relatively rare in the list, but more of them could be found when we identify more words as keys. However, there is still a large proportion of aboutness Key words (>70%) in the top 200 Key words.

Table 4. Proportions of aboutness Key words in the studies varied in the use of cut-off rankings

Studies		Lists	Top n	No. of texts	Benchmark	Aboutness (%)	General interpretation
Main study	3a	1	*25*	(100)	(BNC)	96.0	Difference
		2	*50*	(100)	(BNC)	84.0	
		3	*100*	(100)	(BNC)	76.0	
		4	*200*	(100)	(BNC)	72.0	
Cross-check: benchmark	3b	1	*25*	(100)	(IMR)	40.0	No difference
		2	*50*	(100)	(IMR)	44.0	
		3	*100*	(100)	(IMR)	47.0	
		4	*200*	(100)	(IMR)	51.5	
Cross-check: no. of texts	3c	1	*25*	(60)	(BNC)	100.0	Difference
		2	*50*	(60)	(BNC)	84.8	
		3	*100*	(60)	(BNC)	71.5	
		4	*200*	(60)	(BNC)	70.4	

In the 3b study, IMR was used as the comparative corpus instead of the BNC, and this condition produced results with no difference in the number of aboutness words in the top 25, 50, 100, or 200 lists (approximately 40-51.5%). When we changed the other choice, the number of texts, from 100 to 60 in study 3c, proportion of aboutness held the same pattern as in the main study (3a), whereas fewer aboutness Key words were identified when analyzing more Key words down through the list.

These three studies generally show that cut-off rankings may or may not affect aboutness, depending on the comparative corpus used in an analysis. In a condition using a more specialized corpus (e.g. IMR), different cut-off has no effect on the results in terms of proportion of aboutness words, whereas the use of a general corpus (e.g. BNC) provides a large number of aboutness words. According to Table 4, top 25 Key words are adequate to serve as important words being the focus for further detailed analysis.

The results from this study suggest that in a condition where a target corpus is compared against a more specialized corpus (i.e. IMR), researchers should pay attention to more Key

words than in the condition where a general corpus is used (i.e. BNC). Compared with a more specialized comparative corpus, only 51.5 percent of aboutness Key words were identified, although we considered a total of 200 words. (This is not the case when a different comparative corpus is used. This low proportion of aboutness words may be caused by a key difference between D and IMR, which is more about style and function rather than the content, because they were drawn from the same RAs with the same topics. It could be said that the number of aboutness words should be higher than the results shown in this study if a target and a comparative corpus were collected from different populations of texts, having different content.) If the focus of research is on aboutness Key words, and a general corpus is used as a reference, top 25 is adequate for further detailed analysis of Key words, e.g. by concordancing some interesting Key words to investigate their use in context. However, researchers may need to focus on a greater numbers of Key words if they are interested in investigating other words related to style or function. Based on the results of this research, aboutness Key words are still greater than 70 percent even though we identified 200 words in total.

4. Conclusion

It is known among researchers using the keyword analysis method that different criteria used for identifying Key words will produce different results. However, there is a question of how these different criteria affect the results. This research, then, revealed the extent to which and how different choices of the three main factors: the number of texts in a target corpus, type of benchmark, and cut-off rank, affect Key word results.

Overall, of the three observed factors, the benchmark shows the largest effect on the results of keyword analysis, on both types of Key words and number of aboutness Key words. The number of texts in a target corpus, the second factor, does not show a real effect on the quantity of aboutness Key words. Finally, different cut-off rankings or top n words influence the number of aboutness Key words under the investigated conditions.

When focusing on individual factors, one major finding emerging from the studies varying in the number of texts in the target corpus is that the number of texts in the target corpus does not have a real effect on the Key word results. Proportions of aboutness Key words do not differ even though the target corpora in different simulation studies differed in size or had different numbers of texts. Second, the results from the studies varying in type of benchmark show that specific types of benchmark not only influence proportions of aboutness Key words, but also produce different types of Key words reflecting different aspects of the target corpus. In terms of the proportions of aboutness Key words, the use of a benchmark more similar to the target corpus (i.e. comparing introduction corpus to discussion) is likely to produce fewer aboutness Key words than using a more different benchmark (i.e. the BNC). In terms of the types of Key words, comparison with a specialized corpus produces the most Key words that are specialized terms, whereas comparison with the BNC as a benchmark produces Key words that are related to ELT. Third, the results for cut-off rankings depend on the type of benchmark used in each analysis. According to the findings of this study, top 25 Key words seem to be adequate to serve as a focus for further detailed analysis.

So, how do these findings help add to our understanding of the keyword analysis method? The results of the present study suggest that it is possible to use a small corpus, with a minimum number of 20 texts. From all of the choices under investigation, 20 texts fairly provide a desirable number of aboutness Key words on the top of a keyword list, without significant differences in types of Key words. Moreover, the results suggest that comparing a target corpus with the BNC generates more aboutness Key words than with a more specialized corpus when observing the same top n. Therefore, we recommend that researchers pay particular attention to words in lower rankings when a specialized corpus is used as a

benchmark, to obtain a desirable proportion of aboutness Key words. Finally, when the benchmark used is large and general, the focus can be just the top 25 Key words.

Since the results of this study were derived from a single set of data (the research articles corpus), generalizability of these guidelines would require further investigation. However, due to the fact that different choices affect Key words, these guidelines help ease the decisions we have to make when conducting keyword analysis.

References

Bondi, M., & Scott, M. (Eds.). (2010). *Keyness in texts*. Amsterdam: John Benjamins.

Bruce, I. (2009). Results sections in sociology and organic chemistry articles: A genre analysis. *English for Specific Purposes, 28*(2), 105-124.

Carreon, J. R., Watson Todd, R., & Knox, J. S. (2013). Medical tourism communication of a Thai private hospital website. *Journal of Applied Linguistics and Professional Practice, 8*(2), 165-185.

Crawford, L., Pollack, J., & England, D. (2006). Uncovering the trends in project management: Journal emphases over the last 10 years. *International Journal of Project Management, 24*, 175-184.

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

ElMalik, A. T., & Nesi, H. (2008). Publishing research in a second language: The case of Sudanese contributors to international medical journals. *Journal of English for Academic Purposes, 7*(2), 87-96.

Freddi, M. (2005). Arguing linguistics: Corpus investigation of one functional variety of academic discourse. *Journal of English for Academic Purposes, 4*(1), 5-26.

Gerbig, A. (2010). Key words and key phrases in a corpus of travel writing. In M. Bondi & M. Scott (Eds.)., *Keyness in texts*, 147–168. Amsterdam: John Benjamins.

Jimarkon, P. & Watson Todd, R. (2013). Red or yellow, peace or war: Agonism and antagonism in online discussion during the 2010 political unrest in Thailand. In A. De Rycker & Z. Mohd Don (Eds.)., *Discourse and crisis: Critical perspectives*, 301–322. Amsterdam: John Benjamins.

Kang, N. & Yu, Q. (2011). Corpus-based stylistic analysis of tourism English. *Journal of Language Teaching and Research, 2*(1), 129–136.

Kanoksilapatham, B. (2007). Writing scientific research articles in Thai and English: Similarities and differences. *Silpakorn University International Journal, 7*, 172-203.

Kanoksilapatham, B. (2012). Facilitating scholarly publication: Genre characteristics of English research article Introductions and Methods. *3L: Language, Linguistics and Literature, The Southeast Asian Journal of English Language Studies, 18*(4), 5-19.

Kotzé, E. F. (2010). Author identification from opposing perspectives in forensic linguistics. *Southern Africa Linguistics and Applied Language Studies, 28*(2), 185-197.

Kuteeva, M. & McGrath, L. (2013). The theoretical research article as a reflection of disciplinary practices: The case of pure mathematics. *Applied Linguistics, 36*(2), 215-235.

Kwary, D. A. (2011). A hybrid method for determining technical vocabulary. *System, 39*(2), 175-185.

Paquot, M. & Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. *Language and Computers, 68(1),* 247-269.

Pojanapunya, P. & Watson Todd, R. (2016). Keywords as a clue to disciplinary identities in the discussion sections of research articles. Proceedings of "The 1st RMUTT

International Conference on English Language Teaching (RtICELT) 2016", Rajamangala University of Technology Thanyaburi (RMUTT).

Rayson, P. (2013). Corpus analysis of key words. In C. A. Chapelle (Ed.)., *The encyclopaedia of applied linguistics*, 1–7. Oxford: Wiley-Blackwell.

Scott, M. & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Scott, M. (1997). PC analysis of key words – and key key words. *System, 25*(2), 233-245.

Scott, M. (2009). In search of a bad reference corpus. In D. Archer (Ed.)., *What's in a wordlist? Investigating word frequency and keyword extraction*, 79-91. Aldershot: Ashgate.

Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails.... In M. Bondi & M. Scott (Eds.)., *Keyness in texts*, 43-57. Amsterdam: John Benjamins.

Seale, C. (2008). Mapping the field of medical sociology: A comparative analysis of journals. *Sociology of Health & Illness, 30*(5), 677-695.

Seale, C., Ziebland, S., & Charteris-Black, J. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science and Medicine, 62*(10), 2577-2590.

Wannaruk, A. & Amnuai, W. (2015). A comparison of rhetorical move structure of applied linguistics research articles published in international and national Thai journals. *RELC Journal, 47*(2), 193-211.