

## A Corpus Based Study of Compound Nouns of Society Word Family in Sociology Academic Articles

Suwitchan Un-udom

## บทคัดย่อ

งานวิจัยขึ้นนี้มีจุดประสงค์เพื่อ 1) ศึกษาคำศัพท์ในวงศัพท์ของคำว่า Society ที่ใช้ในการสร้างคำนามประสมบอยที่สุดในบทความที่ตีพิมพ์ในวารสารด้านสังคมวิทยา และ 2) ศึกษาคำประสมของคำในวงศัพท์ที่พบบ่อยในบทความที่ตีพิมพ์ในวารสารด้านสังคมวิทยา ผู้วิจัยใช้วิธีการภาษาศาสตร์คลังข้อมูล (Corpus Linguistics) ในการเก็บข้อมูลแล้วเคราะห์ข้อมูล โดยคลังข้อมูลภาษาประกอบด้วยหน่วยคำจำนวน 350,328 โทเคน (Tokens) จากบทความวิชาการทั้งหมด 48 เรื่อง ผลการศึกษาพบว่าคำว่า “Social” เป็นคำที่ถูกใช้เพื่อสร้างคำนามประสมมากที่สุดเมื่อเทียบกับคำในวงศัพท์ “Society” อีก ๑ นอกร้านนี้ผลของการศึกษาได้เสนอชุดคำ (Word List) ที่รวมคำนามประสมของคำว่า “Social” ที่ถูกใช้มากที่สุดในวารสารด้านสังคมวิทยา การศึกษาแสดงให้เห็นรูปแบบการสร้างคำประสมกับคำสำคัญในบริบทและเป็นแนวทางในการศึกษาคำศัพท์ทั้งในทางด้านการวิจัยและการเรียนการสอน

**คำสำคัญ:** คำสำคัญในบริบท สังคมวิทยา ภาษาศาสตร์คลังข้อมูล การศึกษาคำศัพท์

## Abstract

The current study was conducted with the purpose to identify 1) the member of the society word family that is most frequently used to form compound nouns in the sociology academic articles and 2) the frequent compound nouns containing members of the society word family. The method of corpus-based study was employed to collect and analyze the result of the study. A corpora of sociology articles consisting of 350,328 tokens was compiled from 48 articles in a major sociology publication. The results of the study showed that the word “Social” was the most used member of the word family to create compound nouns in the corpora while a list of compound nouns with the word “Social” was presented. The study suggests compound nouns with the keyword in context in the sociology area and provides academic and pedagogical implication in vocabulary study.

**Keywords:** Keyword in context, Sociology, Corpus linguistics, Vocabulary study

## Introduction

According to Wilkins (1972), “while without grammar very little can be conveyed, without vocabulary *nothing* can be conveyed” (pp. 111–112). It is accepted that vocabulary instruction becomes a core principle in teaching second language. In order to gain more vocabulary knowledge, learners; especially in English for specific purpose (ESP) courses; who have to deal with both content and a specific set of vocabulary, need to have techniques that effectively help to retain lexical items.

One of the important components of vocabulary learning in ESP courses is the instruction of key words in context (Rodriguez & Sadoski, 2000). According to Durant (2008), keywords could be defined as words that are actively used in a specific field. Specifically, keywords should be currently used, polysemous, categorical, actively contested, and part of a cluster. Learning of key words and inflection of the words leads to comprehension of content in the specific areas. Moreover, keywords in a specific context could lead to better techniques to recall the information

(Atkinson&Raugh, 1975). Studies put the efforts to investigate key words in context of specific fields in order to identify words that occur as the key terms of the fields, for instance chemistry (Valipouri&Nassaji, 2013), political science (Pearce, 2014), and medicine (Wang, Liang, & Ge, 2008).

However, it is important to note that key terms in context do not always occur individually. The words could occur with others to form noun phases or compound nouns. Therefore, the study of collocating words with keywords in context might lead to the extension of vocabulary knowledge as learners might gain more lexical items from the compound words consisting of the keywords in context. The existence of phases and compounds containing the key words and collocations has also led to investigations in several areas e.g., engineering (Ward, 2007), agriculture, (Jia-ying, & Yuan-yuan, 2014) and law (Csomay & Petrovic, 2012). In these cases, the study of collocation of keywords in specific fields could be beneficial for both learners and teachers in contributing vocabulary learning.

Corpus-based studies have emerged as a systematic method to examine patterns of language use based on the collection of real use of language. The patterns could be from the level of lexemes, words, compounds, and phases to the holistic level such as genres and styles of registers. Therefore, the method has established itself as the influential tool in language study (Biber et al., 2004). According to Gardner (2007), the capability to examine the vocabulary item using technology support of corpus-based methods could lead to the reliable outcome of the vocabulary patterns in the naturally occurring context.

## Literature Review

### Collocation

Collocation is one of main areas of investigation of the studies using corpus-based methods. The study of words that are located together could lead to deeper understanding of linguistic patterns in the context (Kennedy, 2000). According to Heid and

Gouws (2006) "Collocations are lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in a direct syntactic relation with each other" (pp. 980). By that definition, the collocation of words could be investigated in order to label linguistic features such as grammar, compounds, noun phases, etc.

### Related Studies

Due to the importance of key words in context and the fact that they are not always used in the individual form, researchers attempt to study the co-occurrence of words using corpus linguistic as the method of investigation. In detail, the study of collocations could lead to benefits in vocabulary learning in several aspects.

#### *The study of collocation and technical words*

Co-occurrence of words in the specific field sometimes leads to technical phases and compound nouns of the specific field. Research of corpus-based studies have been conducted to identify technical phases in a field such as Engineering (Ward, 2007), weather forecast (Paegpho&Vitayapirak, 2007), and Guitar rock practice (Impong&Vitayapirak, 2015). For example, Paegpho and Vitayapirak(2011) studied the technical vocabulary and phases in weather forecast and present a sample dictionary. In their work, a corpus of 555,818 running words was collected from the weather forecast texts. It was found that words when occur as a concordance could be considered as technical. The study also resulted in a sample dictionary consisting of technical phases, such as *Cumulonimbus cloud*.

#### *The study of collocation and keywords in context*

The study of collocation can also lead to phases with the key words in context. As mentioned earlier, the study of keywords in context could lead to better understanding of the content of the field. Moreover, the collocation of the keywords could benefit learners in gaining vocabulary size as the collocation of key terms is likely to retain easier

(Rodridguez & Sadoski, 2000). Consequently, studies have been conducted to identify collocation of keywords using corpus-based methods (e.g., Kast-Aigner, 2009; Jia-ying, & Yuan-yuan, 2014; Menon & Makudan, 2012). For instance, Jia-ying, & Yuan-yuan, (2014) studied the collocation of the word *soil* in agricultural journals. The corpus was compiled from the *European Journal of Soil Biology*. The result of the study reveals the collocations of the word *soil* with several parts of speech including nouns, adjective, verbs, and function words as well as technical phrases occurring as a result of collocation of the key word such as “soil temperature”, “physical soil” and “soil fauna”.

### This study

The studies of key words in specific areas could lead to in-depth understanding of language patterns in the fields. The corpus-based study has been established as a method to examine the pattern of vocabulary items. Moreover, collocation of keywords in context could lead to the exploring of the linguistic patterns in the area. Studies have been conducted to investigate the collocation of keywords in context in several areas.

The current study explored to the collocation of the *society* word family in academic articles related to sociology. The corpus study was employed as the method of data collection. The data were collected and analyzed to answer the following research questions.

What member of the *society* word family is most frequently used to form compound nouns in academic sociology articles?

What are the frequent compound nouns containing the most frequent member of the *society* word family?

### Methodology

#### Corpus

The corpus used in this study was compiled from 48 articles published in *The International Journal*

of *Sociology and Policy* in a year (volume 35 issues 5-6 to volume 36 issue 3-4). The journal is in the area of sociology and has published research on sociology and policy for more than 30 years. The corpus consists of 350,328 tokens in total. Since the aim of the study is to investigate the use of language in journal writing, the components which are not related to the context including author’s information, journal information, page number, acknowledgement, and references were excluded.

#### Target words

The study was conducted to investigate the patterns of compound nouns in the sociology articles. The “society” word family was selected from the key words in context of sociology (William, 1983). The word family was selected in this study because it was the only word family that contains all of the members in key words in context. Therefore, all members of the *society* word family including *society*, *social*, *socialize*, *sociology*, *socialism*, *socialist*, *sociological*, *sociologist*, and *societal* (British Council, 2016) were used in the data analysis. The word *socially* was excluded according to the purpose of the study to find compound nouns as adverbs are likely used to create compound verbs.

#### Procedures

After the corpus was compiled, the articles were transformed in to .txt files and analyzed by the software Anconc 3.4.4. The software is a free program published by Laurence Anthony. It is capable of analyzing collocations of the target words and providing concordancing lines for further investigation. All of the target words were analyzed to find the member of the word family that was most frequently used in compound nouns to answer the first research question. For the second research question, the compound nouns consisting of words from the first part of the data analysis were examined in detail to identify the frequently used compound nouns in the articles. The criterion to identify word frequency were those of Hyland and Tse’s (2007); that is lexical items were considered as frequent when they occurred above

the mean of the total number all compound nouns. To identify compound nouns, the definition provided by Nakagawa and Mori (2003) was employed. Specifically, compound nouns could be defined as nouns made with two or more words which could be with open space (e.g., social worker), hyphenated (e.g., corpus-based study), and closed or solid space (work-house). Therefore, the collocation of the word in the “society” word family must meet the criteria below.

- 1) The collocation must form a compound noun.
- 2) The member of “society” word family could be either a head or determiner of the noun.
- 3) The plural and singular forms of a compound word were counted as one compound noun

(e.g., social worker and social workers were counted as one compound).

- 4) Particles of one member were counted as one (e.g., socialized and socializing were counted as one member of word family).

## Results & Discussion

### Most Frequently Used Member to Form Compound Nouns

The first purpose of the current study was to investigate the member of *society* word family that was most frequently used to form compound nouns in the sociology academic article. The analysis from the corpus and concordance can be seen in the table 1.

**Table 1.** The analysis of compound nouns in sociology articles

Members of word family	NO. of compound nouns (CN)	Frequency (n)	Coverage (%)
<b>society</b>	40	74	3.73 %
<b>social*</b>	252	1770	89.12 %
<b>socialize</b>	0	0	0.00 %
<b>sociology</b>	9	21	1.06 %
<b>socialism</b>	5	7	0.35 %
<b>socialist</b>	23	34	1.71 %
<b>sociological</b>	28	45	2.27 %
<b>sociologist</b>	3	3	0.15 %
<b>societal</b>	23	32	1.61 %
<b>Total</b>	383	1986	100 %

As can be seen from the table, it was found that 383 compound nouns was formed with the member of *society* word family. The analysis of An-conc program indicated that these compound nouns occurred 1986 times in the text. In detail, the word *social* was found to be the member that was most frequently used to form compound nouns since 252 items were found having *social* as the modifier of the compounds. Moreover, of all the compound nouns

generated from the member of *society* word family, the compound noun of *social* cover 89.12 %. Therefore, the word *social* is the answer to the first research question.

The result of the study led to an investigation of the use of adjectives in forming compounds or noun phrases. The results indicated that adjectives in the society word family were most frequently used to form compound nouns. The result of this study are

related to Miller (2002) who suggests that modifying nouns by using an adjective is one of the main syntactic characteristics of English. Similarly, Quirk (1985) also indicated that the heads of phrases should be modified by adjectives or adverbs. Finally, Crystal (2003) also suggests that adjectives function mainly as modifiers of nouns. Therefore the greater number of compound nouns formed by the adjectives than by noun and noun are likely to be found.

Moreover, the current result that the word *social* was the most frequent adjective used to form compound nouns could be explained by the contribution of the word in a lexical frequency profile. Vocabprofile at <http://www.lectutor.ca> is a computer program that can identify the most frequent 1000 words of English, the second most frequent thousand

words of English, and Academic words of English (Cobb, 2010). The word *social* is found in the first group of English words while *sociological* and *societal* are rare words because they were not found at all in any of the three lists. Therefore, due to the distribution in English and the main function to modify nouns, the adjective *social* is likely used as the component of most compound nouns found in the sociology articles.

#### *Frequent compound nouns of social*

A further investigation was on the frequent compound nouns of the word *social*. The frequent items were identified by the criteria of Hyland and Tse(2007). That is compound nouns with the word *social* were classified as frequent when they exceed the mean of all the compounds, as shown in the table below.

**Table 2** Most frequent 20 compound nouns of social.

No.	Compound word	Frequency	No	Compound word	Frequency
1	social capital	212	11	social relation	28
2	social exclusion	155	12	social policy	28
3	social assistance	116	13	social trust	27
4	social service	86	14	social protection	21
5	social network	67	15	social care	20
6	social security	63	16	social responsibility	18
7	social worker	54	17	social control	16
8	social insurance	40	18	social groups	16
9	social resource	35	19	social implications	16
10	social justice	34	20	social secretariat	16

Of all 252 compound nouns occurring with *social*, the mean of the all compounds noun was 7.02. Therefore, the items which occur 8 times and above were counted as frequent compound nouns. 42 compound nouns met this criterion. The list of frequent compound noun with *social* could be seen in appendix A., the table 2 shows the first 20 compound nouns, and the word *social capital* ( $n=212$ ), *social exclusion* ( $n=155$ ), and *social assistance* ( $n=116$ ) occurred more than 100 times. These results are discussed in more

detail below.

The occurrence of these compounds could be related to the issues discussed the sociology because sociology could be defined as a branch of social sciences that aims to investigate human social structure, activity, knowledge, policy, and welfare (Giddens, Duneier, Carr, 2016). Therefore, the frequent existence of *social capital*, *social exclusion* and *social assistance* could be seen in the sociology corpus because the meaning of the compound nouns are related to what

is discussed in the area. It could be seen in the results of the study that the word *social* occurs in great a number of compound nouns. The results support the benefit of keywords in vocabulary learning. For example Gu(2010) suggested that the keywords could be beneficial in making wordlists, which is one of the effective strategies in vocabulary learning.

### Conclusion

The results of the study were 1) the word *social* was the member of the *society* word family that was most frequently used to form compound nouns in the sociology articles, and 42 compound nouns of the word *social* were frequently used in the publication of the field (see Appendix A). The results of the study could be implicated in the study of phase and compound in English. Adjectives are used more frequently to form phrases and compounds than other parts of speech, and the words in the frequently occurring list seem to be more used than the words in the less frequent list. Moreover, a pedagogical implication of this study is that learners need to learn keywords in each context as they can be combined with others to form lexical items that also frequently occur in the field. In addition, the result of the study provided the set of compound nouns that could be taught according to the frequency of occurrence in the field.

The current study had some limitations in the process of data collection. The corpus used in the study was compiled from only one year of an academic journal. This issue might affect the representativeness and the balance of the data. Further studies could compile a corpora from various journals in sociology within a longer period of time. Moreover, the current study investigated only a keyword family in the context of sociology. Further studies could investigate other keywords in the area since different results might be found if other components are included. Lastly, compound nouns investigated in the current study might not cover the co-occurrence of words occurring in the areas. Further studies could extend

the collocation of keywords that contain compound characteristics, for example compound verbs as it might be beneficial for both academic and pedagogical purposes.

### References

Atkinson, R. C. & Raugh, M. R. (1975). An application of the Mnemonic keyword method to the acquisition of a Russian Vocabulary. *Journal of Experimental Psychology, 104*(2), 126-133.

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

British Council. (2016). British Council - Word Family Framework. Retrieved April 24, 2016, from <https://www.learnenglish.org.uk/wff/index.html>

Crystal, D. (2003). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Csomay, E., & Petrović, M. (2012). "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System, 40*(2), 305-315. <http://doi.org/10.1016/j.system.2012.05.004>

Duran, A. (2008). The significance is in the selection': Identifying contemporary keywords. *Critical Quarterly, 48* (4), 1-26.

Gardner, D. (2007). Validating the Construct of Word in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics, 28*(2), 241-265. <http://doi.org/10.1093/applin/amm010>

Giddens, A., Duneier, M., Carr, D., & Appelbaum, R. P. (2016). *Introduction to Sociology*. W. W. Norton, Incorporated.

Gu, Y. (2010). Learning strategies for vocabulary development. *Reflections on English Language Teaching, 9*(2), 105-118.

Heid, U., & Gouws, R. H. (2006). A model for a multi-functional dictionary of collocations. In *Attidel XII Congresso Internazionale di Lessico e Linguistica*.

*sicografia: Torino, 6-9 settembre 2006* (pp. 979-988).

Hyland, K., & Tse, P. (2007). Is There an “Academic Vocabulary”? *TESOL Quarterly*, 41(2), 235–253. <http://doi.org/10.2307/40264352>

Impong, N., & Vitayapirak, J. (2015). Corpus-Based Vocabulary Analysis of Rock Guitar Lessons and a Sample Dictionary. *International Journal of Languages, Literature and Linguistics*, 1(1), 39–44.

Jia-ying, H., & Yuan-yuan, S. (2014). A Corpus-based Study on Collocation and Colligation of “Soil” in Agricultural English. *Journal of Northeast Agricultural University (English Edition)*, 21(4), 86–92. [http://doi.org/10.1016/S1006-8104\(15\)30024-6](http://doi.org/10.1016/S1006-8104(15)30024-6)

Kast-Aigner, J. (2009). Terms in context: A corpus-based analysis of the terminology of the European Union’s development cooperation policy. *Fachsprache*, 3(4), 139–152.

Kennedy, G. (2000). *An introduction to corpus linguistics*. Beijing: Language Teaching and Research Press.

Menon, S. & Makudan, J. (2012). Collocations of High Frequency Noun Keywords in Prescribed Science Textbooks. *International Education Studies*, 5(6), 149–160.

Miller, J. (2002). *An Introduction to English Syntax*. Edinburgh University Press.

Nakagawa, H., & Mori, T. (2003). Automatic term recognition based on statistics of compound nouns and their components. *Terminology*, 9(2), 201–219. <http://doi.org/10.1075/term.9.2.04nak>

Paengpho, S., & Vitayapirak, J. (2011). Weather forecast corpus and a sample dictionary. In *Proceedings of the International Conference* (pp. 88–95). Bangkok.

Pearce, M. (2014). Key Function Words in a Corpus of UK Election Manifestos. *Linguistik Online*, 65(13), 23–44.

Quirk, R. (1985). *A comprehensive grammar of the English language*. London: Longman.

Rodriguez, M. & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning*, 50(2), 385–412.

Valipouri, L. & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *English for Specific Purposes*, 12(4), 248–263.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes*, 27(4), 442–458. <http://doi.org/10.1016/j.esp.2008.05.003>

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, 6(1), 18–35. <http://doi.org/10.1016/j.jeap.2006.10.001>

Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Edward Arnold.

Williams, R. (1983). *Keywords A vocabulary of culture and society* (2nd ed.). New York: Oxford University Press.

Appendix A: List of frequent compound nouns of the word “social” social capital, social exclusion, social assistance, social service, social security, social insurance, social network, social justice, social resource, social worker, social relation, social trust, social policy, social protection, social care, social responsibility, social control, social groups, social implications, social secretariat, social welfare, social work, social affairs, social life, social norm, social relationship, social inclusion, social movement, social structure, social support, social problem, social progress, social change, social costs, social democratic, social astuteness, social group, social mobility, social survey, social institution, social context, social practice