

## The Word List Distribution in Social Science Research Articles

**Sorawut Chanasattru**

Language Institute, Thammasat University  
Bangkok, Thailand

**Supong Tangkiengsirisin**

Language Institute, Thammasat University  
Bangkok, Thailand

### Abstract

This study investigates the distribution and coverage of words in New General Service List (NGSL) and the Academic Word List (AWL) in social science research articles. Sixty-four open-access English social science research articles published in 2013-2015 in the ScienceDirect General category were selected and compiled to the Social Science Corpus (SSC). The AntWordProfiler 1.4.0 was utilized to calculate the frequency and coverage percentage of words from the two word lists. Word families in level 1 and level 2 of the NGSL were utilized over 70 percent, whilst level 3 word families were used around 60 percent of the entire SSC. Similarly, 99.65 percent of the AWL word families were discovered. Regarding coverage, the NGSL word families accounted for over 70 percent and the AWL word families covered around 14 percent revealing significant coverage of both word lists. The top 10 NGSL word families represented journals subject areas from which they were derived, whilst the top 10 AWL word families were used more repeatedly and linked with social science research areas. The finding of high distributions and coverage corroborated that the NGSL and the AWL significantly contribute to vocabulary pedagogy in preparing students for reading and writing social science research articles. Additionally, some pedagogical implication guidelines of the NGSL and the AWL such as flash cards, quizzes, and written tests were also introduced.

**Keywords:** academic word list, new general service list, word list coverage, word list implication

**Cite as:** Chanasattru, S., & Tangkiengsirisin, S. (2017). The Word List Distribution in Social Science Research Articles. *Arab World English Journal*, 8 (4).

DOI: <https://dx.doi.org/10.24093/awej/vol8no4.28>

## Introduction

Vocabulary is recognized as one of predicaments for native-speaker and nonnative-speaker learners in academic reading and writing (Shaw, 1991). In Thailand, students have been taught English from kindergarten to university, or for more than twelve years. Nonetheless, insufficient vocabulary knowledge is one of the major difficulties that Thai students face when reading (Aegpongpaow, 2008). The association of vocabulary knowledge and reading comprehension has been exposed in several studies. If students have insufficient vocabulary knowledge or have comprehension of what is being read, they cannot easily learn from context (Nation & Waring, 1997). Students who have high prior vocabulary knowledge level are likely to read quicker than those who have lower or limited knowledge (Calvo, Estevez, & Dowens, 2003). In other words, the better the students' vocabulary knowledge is, the better they achieve reading comprehension (Constantinescu, 2007), English listening tests and other English standardized tests (Nation & Meara, 2002). Consequently, vocabulary knowledge can help reading comprehension, which in turn contributes to the increase of vocabulary size (Chall, 1987).

Numerous studies about the vocabulary size of native English speakers and second language (L2) learners have been conducted. Native English speakers know approximately 20,000 word families; they increase their vocabulary size by 1,000 word families on average each year (Nation & Waring, 1997). A university student knows around 20,000 word families (Goulden, Nation, & Read, 1990; Hirsh & Nation, 1992) excluding proper names, compound words, abbreviations, and foreign words. Likewise, second language (L2) learners needed a substantial vocabulary size to accomplish reading comprehension (Groot, 2000). Second language (L2) learners require knowing at least 3,000 high frequency English words, and their understanding would increase to at least 95 percent coverage of a text (Na & Nation, as cited in Nation & Waring, 1997). This finding was consistent with Hirsh and Nation (1992) that knowledge of around 3,000 - 5,000 word families is required for basic comprehension of a text. However, the disparities of word size of L2 had been reported. Nation (2006) discovered that the required vocabulary size for L2 in reading for pleasure of short or unsimplified novels was 98 percent accounting for 8,000 - 9,000 word families (Nation, 2006). Similarly, the study conducted by Hazenberg and Hulstijn (1996) also revealed that 10,000 base words were the minimal vocabulary size required for university students to comprehend academic text.

One problem in teaching new vocabulary is that teachers cannot decide which vocabulary should be introduced to students because teachers do not know which words frequently appear and are truly representative. Nowadays, powerful personal computers, faster internet speed connections, larger media storage, inexpensive scanners, and intelligent concordance programs greatly contribute to personal corpus and word list creation. Instructors, for example, can easily assemble their own mini or specialized corpora from websites, scan textbooks to digital text files, or access remote electronic databases together with concordance software allowing instructors to compile their own word lists based on corpus in a specified genre

Word lists provide a multitude of advantages in setting up vocabulary learning targets, vocabulary knowledge evaluation, analyzing text difficulty, creating and modifying reading materials, designing vocabulary teaching and learning tools, as well as forming vocabulary courses (Gardner & Davies, 2013). Regarding vocabulary pedagogy, word lists play an essential role in

lexical studies that aim to provide information on the frequency with which words occur in particular genres. Word lists can facilitate teachers or course designers to determine whether these words are meaningful for their students or if they should be overlooked. Likewise, course designers should have word lists to refer to when selecting vocabulary for a language course matching with student's level.

### **Research Questions**

This study investigates the distribution and the coverage of high frequency general English vocabulary from the New General Service List (NGSL) and academic vocabulary from the Academic Word List (AWL) in the social science research articles collected from ScienceDirect website. The result from this study can identify whether words from the NGSL and the AWL are appropriate for vocabulary pedagogy or not. The research question is addressed below:

To what extent do words in the NGSL and the AWL appear on social science research papers?

### **Literature Review**

#### *Word Frequency Lists*

A word frequency list, or a word list, is a list usually sorted either in frequency order or in alphabetical order of all words in a particular corpus together with the frequency of word occurrence in the corpus (Hunston, 2006). Word frequency lists can be generated from a corpus by using concordance software that thoroughly searches every word in that corpus to find out how many tokens and how many different word types there are. The software then generates a frequency list, which can be displayed in rank order of frequency or in alphabetical order (Evison, 2010). Word lists play a crucial role in lexical studies aiming to provide information on the frequency with which words occur in particular areas. The generated word frequency lists can facilitate teachers or course designers to determine whether these words are meaningful for their students or if they should be overlooked.

#### *Previous Studies on Word List Distribution*

The General Service List (GSL) (West, 1953) and the Academic Word List (AWL) (Coxhead, 2000) have been primarily employed in the field of corpus linguistics to investigate the distribution of words in diverse registers. The GSL has contributed to the examination of 2,000 high frequency general English words in specialized corpora. Likewise, the AWL is utilized to identify how often academic vocabulary occurred in academic journals, textbooks, course materials, as well as news. The investigations of lexical distribution have been performed in numerous aspects. Research articles in diverse registers have been studied, for instance, in Medical Science (Chen & Ge, 2007), Agriculture (Martínez, Beck, & Panza, 2009; Muñoz, 2015), Linguistics (Vongpumivitch, Huang, & Chang, 2009), Education (Mozaffari & Moini, 2014), English for Specific Purposes (Shabani & Tazik, 2014), Nursing (Yang, 2015), or Environmental Science (Liu & Han, 2015). Regarding textbooks and course materials, a number of researchers have carried out ranging from primary English course books (Mármol, 2009) to particular course materials used in undergraduate and graduate classes in a variety of subjects: engineering (Chanchanglek, 2010; Kakaew, 2013; Ward, 2009;); maritime navigation (Sun, 2010), culinary courses (Nordin, Stapa, & Darus, 2013), psychology (Kaewphanngam, 2002), or business

computing (Pattaradej, 2009). Moreover, vocabulary occurrence in different genres of online news has been also studied, for example political news (Nakprakhon, 2005), economic news (Manitayakul, 2007), and business news (Liangpanit, 2010).

Concerning coverage, Coxhead and Byrd (2007) indicate that words in the AWL account for approximately 10 percent of academic texts. Many study results corroborated the Coxhead and Byrd's assert. Just to name a few, vocabulary in the AWL accounted for 10.07 percent of medical research articles (Chen & Ge, 2007) and 9.06 percent of agricultural research articles (Martínez, Beck, & Panza, 2009). Likewise, the coverage of AWL was 10.46 percent of the vocabulary in financial corpora (Li & Qian, 2010), whereas vocabulary featured in the AWL covered 12.36 percent of the words in the Asian EFL journal and 17.43 percent of those utilized in the Asian ESP journal (Shabani & Tazik, 2014). In contrast, a few research findings contradicted Coxhead and Byrd's. Mozaffari and Moini (2014), for instance, found that words in the AWL only accounted for 4.94 percent of the vocabulary from education research articles. Similarly, Nakprakhon (2005) and Manitayakul (2007) discovered relatively low coverage of AWL at 5.7 percent in political news corpus and only 3.47 percent in an economic news corpus compiled from the online news agency respectively.

Most researchers examined the frequency and distribution of specialized corpora based on either Coxhead's Academic Word List (AWL) or West's General Service List (GSL). However, the West's GSL was disparaged for being based on out-of-date corpus, for creating a small running word corpus (approximately 2.5 million words) compared to modern standards, and for not clearly defining "word" (Browne, 2014). The New General Service List (NGSL) was created by updating West's General Service List (GSL) based on a 273 million-word corpus, which is about ten times larger than the corpus used in GSL compilation. Nowadays, a tiny number of works apply the NGSL to investigate the frequency and distribution of words in a specialized corpus. To fill this gap, the New General Service List (NGSL) and Coxhead's Academic Word List (AWL) were employed together to investigate the occurrence of frequency and distribution of words in social science research articles gathering from ScienceDirect website.

## Methodology

### Reference Word Lists

#### *The New General Service List (NGSL)*

The NGSL contains 2,801 headwords and 8,456 word family members divided into three different levels of occurrence. Level 1 represents the 1,000 word families that are used most frequently in general English. The second and third levels denote the next most frequent 1,000 and 801 word families, which are found in general English, respectively. The lemmatization of NGSL headwords and word family members were rearranged in a format, as suggested by Paul Nation's Range program, and Laurence Anthony's AntWordProfiler help file so that AntWordProfiler, can identify headwords and their family members characterized by the Bauer and Nation's Level 6 scale (Bauer & Nation, 1993).

#### *The Academic Word List (AWL)*

The AWL was developed by Averil Coxhead, by analyzing the distribution of words not contained in the 2,000 high frequency words from West's GSL, covering 28 subject areas under

four major disciplines – art, science, commerce and law (Coxhead, 2000). The AWL consists of 570 headwords and 3,106 word family members which comprised of word stems and family members characterized by the Bauer and Nation's Level 6 scale (Bauer & Nation, 1993). The AWL file was formatted as suggested by Paul Nation's Range program.

#### Concordancers

The AntWordProfiler was utilized to compare a text corpus against target reference lists of vocabulary based on the research of Paul Nation (Nation & Heatley, 2007) to investigate what words in the text corpus are, or are not presented in word lists, as well as to calculate percentages, frequencies, ranges, and coverage of reference word lists against corpora.

#### The Social Science Corpus (SSC) Compilation

##### *Journal and paper selection*

English social science research articles in the general category on the ScienceDirect website (<http://www.sciencedirect.com>) were collected by using the filter function to search only articles published in open-access journals. The SSC consists of 11 journals published during 2013-2015 to shun selecting outdated words. Each journal was labeled with a unique ID no. from S\_1 to S\_11. These articles were statistically selected by a simple random sampling method. The foreword, editorial, and book reviews were excluded from the study.

##### *Data Standardization*

As suggested by Chen and Ge (2007), research article files were standardized by deleting all graphs, charts, diagrams, equations, bibliographies, text headers and footnotes, author's name and affiliates, or other parts of the articles that concordancer cannot processed. Standardized PDF article files were then converted to UTF-8 plain text format to avoid problems from non-English language word conversion. Next, typographical errors from file conversion were examined. All hyphenated words at the end of columns were joined as one word. After the text files were error-checked, the running words were counted, prior to merging them together, to create a sub-corpus of each journal. All 11 journals contained balanced running words to avoid word bias from subject areas. On average, the number of articles drawn from each journal was between four and eleven, containing approximately 36,000 to 39,000 running words. The profile of each journal applied in the SSC compilation is presented in Appendix A.

##### *Data Processing*

All sub-corpora were loaded into AntWordProfiler 1.4.0 (Anthony, 2014) to generate the vocabulary coverage, range, and the frequency of words in the NGSL, the AWL, and words that are not present in the two word lists called NL.

## **Results and Discussion**

*The Distribution of words in the New General Service List (NGSL) and in the Academic Word List (AWL) in Social Science Research Articles*

**Table 1.** *The Distribution and coverage of the NGSL and the AWL coverage in Social Science Corpus (SSC)*

Reference Word List	Number of Word Family	Word Family Occurred	Word Family (%)	Type	Type (%)	Running Words	Running Words (%)
NGSL	Level 1 (1000 Word Families)	888	88.80%	2,249	12.54%	273,236	65.91%
	Level 2 (1000 Word Families)	716	71.60%	1,372	7.65%	21,453	5.18%
	Level 3 (801 Word Families)	510	63.67%	853	4.76%	8,736	2.11%
AWL	570 Word Families	568	99.65%	2,216	12.36%	57,445	13.86%
Not in Both Word Lists (NL)		N/A	N/A	11,244	62.70%	53,675	12.94%
<b>Total</b>		<b>2,682</b>	<b>-</b>	<b>17,934</b>	<b>100.00%</b>	<b>414,545</b>	<b>100.00%</b>

Table 1 represents distributions and the coverage of the NGSL and the AWL word families in the Social Science Corpus (SSC). To begin with word distributions, word families in the NGSL and the AWL covered 2,682 words altogether in the SSC. Among the first 1,000 words in the NGSL level 1, the total of 888 word families (88.80 percent) were discovered, whereas the figures for the next 1,000 word families in level 2 and the next 801 word families in the following level were 716 (71.60 percent) and 510 (63.67 percent) word families, respectively. Of the 570 word families of the AWL, 568 word families (99.65 percent) were found in the SSC. Out of the 568 word families, only *levy* and *rigid* were excluded from the SSC because these two word families may be rarely employed in the social science field, but may be used commonly in finance or tax field instead.

In regard to coverage, word families in the NGSL totally accounted for 73.20 percent in The SSC. The first 1,000 words in level 1 covered 65.91 percent of NGSL, demonstrating that the words in level 1 have a vital role in social science academic articles. Following this, words in levels 2 and 3 accounted for 5.18 percent and 2.11 percent, respectively. Pertaining to academic word lists, 57,445 word families (13.86 percent) were words from the AWL. This coverage corresponded to the Coxherd (2000), Coxhead and Byrd (2007) and several researchers: Chen and Ge (2007), Martínez et al. (2009), Li and Qian (2010), as well as Shabani and Tazik (2014) that the coverage of the AWL was around 10 percent of academic texts. The third category was words which were not in either the NGSL or AWL, known as NL words, covering 53,675 running words (12.94 percent). This is relatively analogous to the coverage percentage attained by the AWL. Regarding word type, NL words were the largest group covering 11,244 types and accounting for 62.7 percent of the SSC. The finding, however, demonstrated that the NGSL covered 73.20 percent of the SSC which went against Browne's assertion that the coverage of the NGSL was at 85.35

percent in English language weekly economics newspaper (Browne, 2014) representing journals in social science.

To elucidate the coverage of the NGSL, the AWL and NL words in each social science journal, the coverage percentage is exhibited in Table 2. As can be seen, each journal contained a relatively similar percentage of the NGSL, the AWL and NL coverage. High frequency words in NGSL covered approximately 68.90 to 74.58 percent of those in all 11 journals. The Studies in Communication Sciences journal (S\_6) exhibited the highest coverage of high frequency words at 75.06 percent. This was followed by coverage of the Sport Management Review (S\_5) at 74.58 percent, the Procedia - Social and Behavioral Sciences (S\_3), and the Journal of Social Studies Research (S\_7), which shared the same percentage of coverage (73.99 percent). In contrast, the lowest coverage of high frequency words in NGSL was found in the Environmental Innovation and Societal Transitions journal, (S\_2) (68.90 percent). Concerning academic words, although the coverage of words in AWL in all journals was comparatively low, at 11.03 to 20.33 percent. As mentioned earlier, this coverage corroborated the findings of Coxhead (2000) and many researchers (Chen & Ge, 2007; Martínez et al., 2009; Li & Qian, 2010; Shabani & Tazik, 2014) that words in the AWL accounted for around 10 percent of academic texts. The Environmental Innovation and Societal Transitions journal (S\_2) demonstrated the greatest extent of coverage of words in AWL, accounting for 20.33 percent. The Energy Research & Social Science journal (S\_1), and the Travel Behaviour and Society journal (S\_9) were found to be next in line in this respect, accounting for 16.95 and 15.60 percent, respectively.

**Table 2.** Coverage of the NGSL and the AWL across 11 journals

ID	Journal Titles	Token	Coverage (%)			NGSL and AWL Coverage (%)	Total Coverage (%)
			NGSL	AWL	NL		
S_1	Energy Research & Social Science	38,131	71.96	16.95	11.09	88.91	100.00
S_2	Environmental Innovation and Societal Transitions	38,261	68.90	20.33	10.77	89.23	100.00
S_3	Procedia - Social and Behavioral Sciences	37,878	73.99	13.00	13.01	86.99	100.00
S_4	Public Relations Review	37,732	72.67	12.42	14.91	85.09	100.00
S_5	Sport Management Review	38,799	74.58	13.49	11.93	88.07	100.00
S_6	Studies in Communication Sciences	35,985	75.06	11.07	13.87	86.13	100.00
S_7	The Journal of Social Studies Research	38,074	73.99	11.03	14.98	85.02	100.00
S_8	The Social Science Journal	38,253	73.95	12.31	13.74	86.26	100.00
S_9	Travel Behaviour and Society	38,869	73.48	15.60	10.92	89.08	100.00
S_10	Wine Economics and Policy	36,481	73.70	12.12	14.18	85.82	100.00
S_11	Urban Climate	36,082	72.95	13.82	13.23	86.77	100.00

---

**Total**                      **414,545**

---

### High Frequency Words in Social Science Research Articles

As presented in Table 3, all top 10 NGSL word families accounted totally for 3.14 percent whilst the coverage of all top 10 AWL word families had three times outnumbered than that of NGSL and covered about one-tenth of the whole corpus. This finding revealed that general English words were used more diverse than academic vocabulary employed in social science academic writing. Clearly, most academic words, for example *Research*, *Policy*, *Analyse*, *Participate*, *Community*, *Respond*, or *Economy* was used more repeatedly and connected with social science genre. Furthermore, the disparity of coverage may stem from higher family members of the AWL.

**Table 3.** *Top 10 NGSL word families in the SSC*

Rank	NGSL Word Families	Frequency	Range	Coverage %	Cumulative Coverage %
1	Study	1510	11	0.50%	0.50%
2	Use	1283	11	0.42%	0.92%
3	Social	973	11	0.32%	1.24%
4	Public	971	11	0.32%	1.56%
5	Also	934	11	0.31%	1.87%
6	Time	863	11	0.28%	2.15%
7	Change	844	11	0.28%	2.43%
8	Model	733	11	0.24%	2.67%
9	High	712	11	0.23%	2.91%
10	Make	691	11	0.23%	3.14%

In contrast, the NGSL word family members only contained words in all inflected forms, and verb in present tense and past tense forms. AWL's word family members, on the other hand, counted words in all inflected forms, some derivational suffixes, and included words with some prefixes as members under the same headwords. For example the headword *Study* in NGSL had only three members containing only inflectional suffixes i.e. *Studies*, *Studied*, and *Studying*, whilst the headword *Research* in AWL consisted of five members including both inflectional and derivational suffixes: *Researched*, *Researcher*, *Researchers*, *Researches*, and *Researching* as presented in Table 4.

**Table 4.** *Top 10 AWL word families in the SSC*

Rank	AWL Word Families	Frequency	Range	Coverage %	Cumulative Coverage %
1	Research	890	11	1.55%	1.55%
2	Policy	805	11	1.40%	2.95%
3	Analyse	777	11	1.35%	4.30%
4	Participate	707	11	1.23%	5.53%
5	Community	680	11	1.18%	6.72%
6	Respond	638	11	1.11%	7.83%
7	Economy	637	11	1.11%	8.94%
8	Project	633	11	1.10%	10.04%



9	Process	628	11	1.09%	11.13%
10	Individual	608	11	1.06%	12.19%

Regarding word dispersion in each journal, most of top 10 NGSL word families can indicate journals subject area from which they were derived. As presented in Table 5, some NGSL top 10 word families in S\_5: *Sport, League, Club*, and *Sponsor* can indicate that these words may come from sport science journals, or some S\_11 top 10 word families: *Climate, Change, Air, Quality*, and *Water* can demonstrate that these journals would relate to climate changes or environment. Similarly, some top 10 word families of S\_7 such as *Teacher, Study, Teach, Education, Student* and *Social* can show that these words would derive from teaching and education. Likewise, some of those in S\_9: *Travel, Trip*, and *Behavior* can point out that these journals were relevant to travel behavior. To sum up the dispersion, the majority of the top 10 word families (58) found only in one journal. However, some top 10 word families can be found more than one journal. The word *study* was highest found in six journals out of eleven whilst the words *social* and *use* were discovered in four journals, as well as the words *change, development, group, market*, and *support* were distributed in three journals (See Appendix Table B1 for more information).

**Table 5.** Top 10 NGSL word family dispersion in each journal

Rank	S_1	S_2	S_3	S_4	S_5	S_6
1	System	System	Study	Public	Sport	Study
2	Carbon	Local	Support	Relation	League	Use
3	Market	Change	Group	History	Relationshi p	Voice
4	Use	Support	Use	Developme nt	Brand	Information
5	Business	Group	Service	Business	Club	Title
6	Social	Carbon	Station	Historical	Drive	Standard
7	Support	Actor	Social	Practice	Developme nt	Expectation
8	Also	Level	Student	Education	Sponsor	Friendship
9	Country	Need	Also	Also	Study	Friend
10	Developme nt	Case	People	State	Market	User
Rank	S_7	S_8	S_9	S_10	S_11	
1	Teacher	Social	Travel	Wine	Climate	
2	Study	Study	Time	Quality	Change	
3	Teach	Case	Trip	Reputation	Barrier	
4	Education	Child	Choice	Product	Air	
5	Student	Marriage	Model	Market	Plan	
6	History	Family	Decision	High	Health	
7	Social	Effect	Probability	Cluster	House	
8	Race	Time	Use	Value	Quality	

9	Experience	Same	Behavior	Price	Water
10	Historical	Group	Change	Production	Study

Unlike the NGSL, the top 10 AWL word families were difficult to identify the subject area from which they were obtained. As presented in Table 6, some word families like *Consume*, *Strategy*, *Professional*, *Respond*, or *Research* cannot easily identify that the S\_5 journal related to sport science comparing to those from NGSL. Regarding word dispersion, top 10 AWL word families exhibited more dispersion than those of NGS. In a nut shell, the majority of the top 10 AWL word families (39) were found only in one journal, whilst 15 word families occurred in two journals, and 9 out of 77 word families distributed more than three journals. The word *analyses* and *research*, for instance, were dispersed in seven journals out of eleven. Similarly, the word *participate* can be found in six journals and the words *individual*, *process*, and *respond* presented in four journals. This result revealed more dispersion of academic words across social science academic articles than those used in general English (See Appendix Table B2 for more information).

**Table 6.** Top 10 AWL word family dispersion in each journal

Rank	S_1	S_2	S_3	S_4	S_5	S_6
1	Energy	Community	Psychology	Corporate	Consume	Communicate
2	Project	Policy	Access	Theory	Strategy	Research
3	Policy	Energy	Research	Ideology	Professional	Respond
4	Finance	Project	Economy	Communication	Respond	Evaluate
5	Design	Transit	Survey	Role	Research	Participate
6	Significant	Innovate	Survive	Focus	Process	Task
7	Sector	Sustain	Individual	Structure	Participate	Process
8	Research	Invest	Process	Professional	Analyse	Text
9	Analyse	Sector	Area	Culture	Involve	Flexible
10	Technology	Analyse	Participate	Analyse	Outcome	Individual

  

Rank	S_7	S_8	S_9	S_10	S_11
1	Economy	Individual	Substitute	Consume	Adapt
2	Participate	Resource	Transport	Vary	Energy
3	Significant	Estimate	Theory	Region	Impact
4	Immigrate	Sex	Individual	Purchase	Estimate
5	Concept	Participate	Research	Research	Participate
6	Identify	Analyse	Survey	Analyse	Identify
7	Research	Data	Vary	Globe	Issue
8	Analyse	Vary	Alternative	Code	Process
9	Data	Tradition	Respond	Create	Region

---

10	Culture	Respond	Utilise	Area	Economy
----	---------	---------	---------	------	---------

---

### Pedagogical Implications of the NGSL and AWL

Owing to significant coverage of the NGSL and the AWL in social science research papers, accounting totally for 87.06 percent, both word lists should be introduced in vocabulary teaching. Initially, teachers can skip function words since these words have little meaning by themselves; they assign grammatical relationships or define function (Meyer, 2009). Instead, teachers should pay attention to content words since they convey meanings of thing, action, state, quality, or event (Nunan, 2013). To familiarize students with high frequency content words in social science articles, learning from word cards or flashcards are useful for both essential high frequency words which need to be introduced in a class room, and low frequency words which can be assigned to students to practice alone (Nation & Meara, 2002). Flashcard method is an effective direct vocabulary learning technique in terms of the return for time and effort. This technique is suitable for learners who do not know many high frequency words or need to increase their vocabulary size since learners can control vocabulary learning repletion and focus on each word that cannot be obtained easily from context or dictionary use (Nation, 2001; Nation & Meara, 2002).

Flashcards can be created by writing a word on one side and putting its definition or other lexical information such as pronunciation, part of speech, grammatical patterns, collocations, or contexts in use in the other side. In addition, mnemonic tools e.g. picture or words in learners' first language (L1) linking its meaning to learning words can effectively increase learners' memory of new vocabulary. Including words and pictures in the same flashcard would enhance learners' understanding and faster learning speed (Kellogg & Howe, 1971). Another tip for learning difficult words is to place these words at the beginning and at the end of the flashcards. This method can contribute to learners' understanding from primacy and recency effects (Baddeley, 1990) that learners are apt to remember the first word since it begins first and likely to remember the last word because it has just already seen.

Assistance from computer technology can facilitate teachers to generate their own flashcards easily without wasting time on pen and paper. For example, Quizlet (<https://quizlet.com>) is a free online vocabulary flashcard generator assisting teachers to create flashcards by uploading the NGSL or the AWL words of their choice. Apart from flashcard creation, the website can generate online games, exercise, and interactive game called Scatter. Memrise (<http://www.memrise.com>) is another online flashcard-learning website with different advantage by adding a Mempty -- a picture or words that associate with the learning word on the flashcard that can increase learners' memory of a target words as proposed by Nation and Meara (2002).

The NGSL and the AWL, furthermore, can also be applied in vocabulary teaching via a test in order to recall words learnt from the flashcard method. The New General Service List Test (NGSLT) (Browne, Culligan, & Phillips, 2014) is a diagnostic test of written receptive knowledge of words in the NGSL. The test contains 100 questions, and each question consists of the target word, presented in bold face, followed by a short sentence containing the target word in a simple non-defining sentence. This sentence assists students to indicate the part of speech of the target word, limit the target word meaning if the target word has more than one possible sense, and

provide a clue for students to predict the meaning of the target word. Sample sentences and the answers are based on the most common meaning or by consultation concordance lines in the Corpus of Contemporary American English. Also, teachers can verify answers from answer keys provided in the website. The example of the NGSL test is presented in Figure 1.

**Test of Written Receptive Knowledge of the New General Service List**

**Level 1**

<p><b>1 charge:</b> They are the charges.</p> <p>a. important things to think about</p> <p>b. prices for a service</p> <p>c. good things</p> <p>d. reasons</p>	<p><b>11 include:</b> We are including it.</p> <p>a. paying</p> <p>b. changing</p> <p>c. adding</p> <p>d. reading</p>
<p><b>2 case:</b> This is a good case.</p> <p>a. place to study</p> <p>b. way something works</p> <p>c. example of something</p> <p>d. plan for the future</p>	<p><b>12 building:</b> Where is the building?</p> <p>a. group of people working together</p> <p>b. road</p> <p>c. middle part</p> <p>d. place to live or work</p>
<p><b>3 different:</b> They are different.</p> <p>a. easy to see</p> <p>b. large</p> <p>c. not easy</p> <p>d. not the same</p>	<p><b>13 true:</b> that is true.</p> <p>a. correct</p> <p>b. different</p> <p>c. interesting</p> <p>d. natural</p>
<p><b>4 room:</b> Where is the room?</p> <p>a. thing we read</p> <p>b. thing to drive</p> <p>c. place to buy things</p> <p>d. space in a building</p>	<p><b>14 teacher:</b> They are teachers.</p> <p>a. people with children</p> <p>b. workers in schools</p> <p>c. leaders in a company</p> <p>d. young people</p>

**Figure 1.** The Example of the NGSL Writing Test. Reprinted from *NGSL/NAWL Test*, by Browne, Culligan, and Phillips, 2014, retrieved from <http://www.newgeneralservicelist.org/ngsl-levels-test/>

Regarding academic word test, various websites offered free exercises based on 570 word families of the AWL. To name a few, Vocabulary Exercises for the Academic Word List (<http://www.englishvocabularyexercises.com/AWL/id21.htm>) provides a number of fill-in-the-blank exercises categorized by the sub-list 1 to sub-list 10 of AWL academic words. AWL Academic Word List Test (<http://www.ecenglish.com/learnenglish/lessons/awl-academic-word-list-test>) allows learners to pick the closest meaning of AWL words via multiple choice questions or cloze tests. Additionally, vocabulary in EAP website (<http://www.uefap.com/vocab/exercise/exercise.htm>) gives learners to practice reading skill via diverse genres of passages with the AWL word highlighted, so that learners can pay attention to academic words.

### Limitations of This Study

The limitation of this study came from the limitations of the AntWordProfiler. The program counted words frequency by grouping words with identical forms that the program can recognize rather than grouping them by their classes or meaning. Furthermore, the SSC was created from academic articles in social science (General) on the ScienceDirect. However, social science in ScienceDirect is composed of 13 specific subject areas: Archaeology, Development, Education, Geography, Planning and Development, Health, Human Factors and Ergonomics, Law, Library

and Information Sciences, Linguistics and Language, Safety Research, Social Sciences (General), Sociology and Political Science, and Transportation. Additionally, the SSC size was around 400,000 running words, and was therefore comparatively smaller than other corpora of Coxhead's AWL, West's GSL, Mozaffari's Academic Words in Education Research Articles, or Yang's Nurse Academic Word List (NAWL).

### Conclusion

Both high frequency English words in the NGSL and academic words in the AWL play a vital role in social science research articles from their high distribution and coverage of words across 11 journals. To begin with word family distribution, 2,114 word families of all three levels of the NGSL and 568 word families in the AWL were discovered in the SSC. Word families in level 1 and level 2 of the NGSL were employed at 88.8 percent and 71.60 percent, respectively, whilst word families in level 3 appeared at a lower rate of 63.67 percent. Regarding notable word coverage, the NGSL accounted for 73.20 percent of the SSC. Similarly, 99.65 percent of AWL academic word families accounted for 13.86 percent which was consistent with the conclusions of Coxhead and several studies. Nearly as many NL words were discovered in the SSC as academic words in the AWL at around 13 percent. The features the top 10 word families from the two word lists were investigated; general English words were used more diversely while most academic words were used more repeatedly and connected with social science genre. Additionally, most top 10 general English words of the NGSL can indicate journals subject areas from which they were derived, while the top 10 academic words of the AWL were hard to point out subject areas. In terms of vocabulary pedagogy, teachers or course designers of social science class can make use of words from the NGSL and the AWL to familiarize learners with words in social science texts. Teachers may introduce word families in the NGSL level 1 because the NGSL level 1 consists of the most frequent words among level 2 and 3. After they are familiar with the word families in level 1, teachers may move to word families from the NGSL level 2, 3, respectively. After students have learnt the word families from the NGSL, teachers then teach word families from the AWL since the coverage of AWL was far lower than the NGSL.

### About the Authors:

**Sorawut Chanasattru** obtained his bachelor's degree in business administration from Kasetsart University and graduated with a master's degree in English for Careers from the Language Institute, Thammasat University.

**Supong Tangkiengsirisin** is an Associate Professor in Applied Linguistics at the Language Institute of Thammasat University. His research interests focus on second language writing, written discourse analysis, and corpus-based analysis.

### References

- Aegpongpaow, O. (2008). A qualitative investigation of metacognitive strategies in Thai students' English academic reading. (Unpublished Master Thesis), *Srinakharinwirot University, Thailand*.
- Anthony, L. (2014). AntWordProfiler (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net>
- Baddeley, A. (1990). *Human memory: Theory and practice*. London: Erlbaum.

- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6, 253-279.
- Browne, C. (2014). The New General Service List Version 1.01: Getting Better All the Time. *Korea TESOL Journal*, 11(1), 35-50
- Browne, C., Culligan, B., & Phillips, J. (2014). NGSL/NAWL Test. Retrieved from <http://www.newgeneralservicelist.org/ngsl-levels-test/>
- Calvo, M. G., Estevez, A., & Dowens, M. G. (2003). Time course of elaborative inferences in reading as a function of prior vocabulary knowledge. *Learning and Instruction*, 13(6), 611-631.
- Chall, J. (1987). Two vocabularies for reading: Recognition and meaning. In M.G. McKeown, and M.E. Curtis (Eds). *The nature of vocabulary acquisition* (pp. 7-17). Hillsdale, NJ: Erlbaum.
- Chanchanglek, S. (2010). *A Corpus Analysis of English Vocabulary Input in Course Materials Used for Engineering Students* (master's thesis). Kasetsart University. Thailand.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, 26(4), 502-514.
- Constantinescu, A. I. (2007). Using technology to assist in vocabulary acquisition and reading comprehension. *The Internet TESL Journal*, 13(2).
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Coxhead, A., & Byrd, P., (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *J. Second Lang. Writing*, 16(3), 129-147.
- Evison, J. (2010). What are the basics of analysing a corpus? In *The Routledge Handbook of Corpus Linguistics*. London, England: Routledge. pp. 122-135.
- Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.
- Goulden, R., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341-363.
- Groot, P. J. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, 4(1), 60-81.
- Hazenbergh, S., & Hulstijn, J.H. (1996) Defining a minimal receptive second-language vocabulary for non-native university students: an empirical investigation. *Applied Linguistics* 17: 145–163.
- Hirsh, D. & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8(2), 689-696.
- Hunston, S. (2006). Corpus linguistics. *Encyclopaedia of Language & Linguistics (Second Edition)*, 234–248. doi:10.1016/B0-08-044854-2/00944-5
- Kaewphanngam, C. (2002). *A Corpus Analysis of Psychology Texts as a Basis for the Development of Teaching Materials in English for Academic Proposes*. M.A. Mahidol. Mahidol University.
- Kakaew, S. (2013). *A corpus-based study of the lexis in marketing* (Unpublished master's thesis). Khon Kaen University, Thailand.
- Kellogg, G.S. & Howe, M.J.A. (1971). Using words and pictures in foreign language learning. *Alberta Journal of Educational Research*, 17, 89-94.

- Liangpanit, C. (2010). *The development of a corpus-based vocabulary package for business English* (Unpublished master's thesis). Suranaree University of Technology, Thailand.
- Li, Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, 38(3), 402-411.
- Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, 39, 1-11.
- Manitayakul, C. (2007). *An Analysis of Academic vocabulary in Economic News* (Unpublished master's thesis). Srinakharinwirot University, Thailand.
- Mármol, G. A. (2009). Vocabulary input in EFL textbooks: frequency levels. In P. Cantos Gómez and A. Sánchez Pérez (Eds.), *A Survey of Corpus-based Research / Panorama de investigaciones basadas en corpus*. Murcia: Asociación Española de Lingüística de Corpus. 771-783.
- Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes* doi:10.1016/j.esp.2009.04.003
- Meyer, C. F. (2009). *Introducing English linguistics*. Cambridge: Cambridge University Press.
- Mozaffari, A., & Moini, R. (2014). Academic Words in Education Research Articles: A Corpus Study. *Procedia-Social and Behavioral Sciences*, 98, 1290-1296.
- Muñoz, V. L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes*, 39, 26-44.
- Nakprakhon, B. (2005). *Academic vocabulary in political reports* (Unpublished master's thesis). Srinakharinwirot University, Thailand.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006) How large a vocabulary is needed for reading and listening? *Canadian Modern Languages Review* 63: 59–82.
- Nation, I. S. P., & Heatley, A. (2007). Range. Retrieved from <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>
- Nation, I.S.P., & Meara, P. (2002). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2<sup>nd</sup> ed., pp. 34-52). London: Arnold.
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, 14, 6-19.
- Nordin, M. N. R., Stapa, S. H., & Darus, S. (2013). Developing a specialized vocabulary word list in a composition culinary course through lecture notes. *Advances in Language and Literary Studies*. 4(1), 78-88.
- Nunan, D. (2013). *What is this thing called language?* (2<sup>nd</sup> ed.). Basingstoke: Palgrave Macmillan.
- Pattaradej, R. (2009). A Corpus-based Analysis of Business Computer Terms. *ASIALEX 2009, The 6<sup>th</sup> International Conference on "Dictionaries in Education"*.
- Shabani, M. B., & Tazik, K. (2014). Coxhead's AWL Across ESP and Asian EFL Journal Research Articles (RAs): A Corpus-based Lexical Study. *Procedia-Social and Behavioral Sciences*, 98, 1722-1728.
- Shaw, P. (1991). Science research students' composing processes. *English for Specific Purposes*, 10(3), 189-206.
- Sun, C. T. (2010). *A corpus-based lexical study of maritime navigational English materials* (Doctoral dissertation, Ming Chuan University).

- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33-41. doi:10.1016/j.esp.2008.08.003
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*. doi:10.1016/j.esp.2009.04.001
- West, M. (1953). *A General Service List of English words*. London, England: Longman, Green & Co.
- Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, 37, 27-38.



**Appendix A****Table A1***Summary of journal applied in the Social Science Corpus (SSC)*

<b>ID</b>	<b>Journal Title</b>	<b>Subject Areas</b>	<b>Article Used</b>	<b>Types</b>	<b>Tokens</b>
S_1	Energy Research & Social Science	Energy Technologies, Fuels, Resources, and Energy Production Affecting People	4	4,501	38,131
S_2	Environmental Innovation and Societal Transitions	Innovations and Socio-Economic Transitions, Environmental Problems, and Environmental Sustainable Economy	4	4,194	38,261
S_3	Procedia - Social and Behavioral Sciences	Social Behavioral Sciences concerning Arts and Humanities	11	4,904	37,878
S_4	Public Relations Review	Public Relations, Mass Communications, Organizational Communications, Marketing Management and Public Policy Formation	6	5,038	37,732
S_5	Sport Management Review	Sport Management and Marketing	4	4,504	38,799
S_6	Studies in Communication Sciences	Public Communication	5	4,489	35,985
S_7	The Journal of Social Studies Research	Social Science Studies	5	4,384	38,074
S_8	The Social Science Journal	Social Sciences, Humanities, and Natural Sciences	7	4,709	38,253
S_9	Travel Behaviour and Society	Travel Behavior, Transportation and Environmental, Transportation Geographic Information Systems	6	4,067	38,869
S_10	Wine Economics and Policy	Wine Business and Economics Management	6	4,880	36,481
S_11	Urban Climate	Urban Climatic Conditions and Change concerning Geography Demographic, and Socioeconomic	6	4,316	36,082
<b>Total</b>			<b>64</b>	<b>49,986</b>	<b>414,545</b>
<b>Average</b>					<b>37,685.91</b>
<b>SD</b>					<b>1,029.38</b>

**Appendix B****Table B1***Summary of Top 10 NGSL word family dispersion across all journals*

Word Dispersion Across Journals	NGSL Word Families (The Total Number of Words)
6	Study (1)
4	Social, Use (2)
3	Also, Change, Development, Group, Market, Support (6)
2	Business, Carbon, Case, Education, Historical, History, Quality, Student, System, Time (10)
1	Actor, Air, Barrier, Behavior, Brand, Child, Choice, Climate, Club, Cluster, Country, Decision, Drive, Effect, Expectation, Experience, Family, Friend, Friendship, Health, High, House, Information, League, Level, Local, Marriage, Model, Need, People, Plan, Practice, Price, Probability, Product, Production, Public, Race, Relation, Relationship, Reputation, Same, Service, Sponsor, Sport, Standard, State, Station, Teach, Teacher, Title, Travel, Trip, User, Value, Voice, Water, Wine (58)

**TABLE B2**

*Summary of Top 10 AWL word family dispersion across all journals*

Word Dispersion Across Journals	AWL Word Families (The Total Number of Words)
7	Analyse, Research (2)
6	Participate (1)
4	Individual, Process, Respond (3)
3	Economy, Energy, Vary (3)
2	Area, Communicate, Consume, Culture, Data, Estimate, Identify, Policy, Professional, Project, Region, Sector, Significant, Survey, Theory (15)
1	Access, Adapt, Alternative, Code, Community, Concept, Corporate, Create, Design, Evaluate, Finance, Flexible, Focus, Globe, Ideology, Immigrate, Impact, Innovate, Invest, Involve, Issue, Outcome, Psychology, Purchase, Resource, Role, Sex, Strategy, Structure, Substitute, Survive, Sustain, Task, Technology, Text, Tradition, Transit, Transport, Utilise (39)