The Academic Article Word List for Social Sciences[1]

Deny A. Kwary[2] and Almira F. Artha[3], Airlangga University, Surabaya, Indonesia

Abstract

The significant contribution of a word list in English language teaching and learning has encouraged researchers to produce various types of word lists. Each word list has its own purpose and is used by a particular group of people. In this article, we created and tested a word list called the Academic Article Word List for Social Sciences (AAWL-SS). This word list is useful for teachers who want to create materials for teaching students of social sciences and researchers who want to publish articles in international journals. In the context of English as a second or foreign language, the word list is useful for translators who translate articles in the social sciences so that the choice of words is more suitable for the international journals. The AAWL-SS was derived from 122 carefully selected open access journal articles. The size of the corpus is 1,040,259 word tokens. The tests show that the AAWL-SS, which only consists of 350 word families, provides a better coverage of the vocabulary used in journal articles, if compared with the 570 word families of the AWL, 880 word families of the NAWL, and the 1000 word families of the BNC-COCA baselist3.

Resumen

La contribución importante de una lista de palabras en la enseñanza y el aprendizaje del inglés ha estimulado a los investigadores para generar varios tipos de listas de palabras. Cada lista tiene su propio propósito y es utilizada por un grupo especial de personas. En este trabajo generamos y probamos una lista de palabras llamada Lista Mundial Académica de Palabras para las Ciencias Sociales. Esta lista de palabras le es útil a los maestros que quieren generar material para enseñar a estudiantes e investigadores de las ciencias sociales que deseen publicar sus trabajos en las revistas internacionales. En el contexto del inglés como segunda lengua o lengua extranjera, la lista de palabras es útil para los traductores que traducen artículos en ciencias sociales por lo que la selección de las palabras sea más adecuada en las revistas internacionales. El AAWL-SS se obtuvo de 122 artículos de revistas de acceso abierto. El tamaño del corpus es de 1040259 "tokens" (número total de palabras en un texto sin importar cuántas veces se repiten). La prueba muestra que el AAWL-SS, que consiste sólo de 350 familias de palabras, proporciona una mejor cobertura del vocabulario utilizado en los artículos de las revistas si se comparan con las 570 familias de palabras del AWL, 889 familias de palabras del NAWL y las 1000 familias de palabras del BNC-COCA baselist3.

Introduction

A number of researchers and teachers have mentioned the significant role of vocabulary in applied linguistics research and in language teaching (Bogaards & Laufer, 2004; Carter, 2012; Nation, 2001). However, given the vast number of words covered in the vocabulary of a language, it is virtually impossible to learn all the words, let alone teach them to second language learners. Consequently, several researchers have created word lists containing the words considered important to be studied by language learners (e.g., Browne, Culligan & Phillips, 2013; Coxhead, 2000; Gardner & Davies, 2014). The creation of word lists develops in line with the advances in the research in corpus linguistics. Sinclair (2005) defines a corpus as a collection of pieces of language text in electronic form, selected to represent a language or language variety. A corpus analysis will inform us how speakers and writers actually use the vocabulary and grammar in a language.

The development of corpus linguistics in the English language began in 1940 with Rudolph Quirk as the pioneer (Aijmer & Altenberg, 1991). Several studies have shown that research in corpus linguistics has played a significant role in the study of language (cf. Campoy, Bellés-Fortuño & Gea-Valor, 2010; Laviosa, 1998). Through corpus

---

linguistics, large amounts of text data can be processed and elaborated effectively and efficiently. Nowadays, a number of researchers have created various types of corpora, particularly those using the English language. Some of the examples are the British Academic Written English (BAWE) corpus (Alsop & Nesi, 2009), the Corpus of Contemporary American English (Davies, 2010) and the Vienna-Oxford International Corpus of English (VOICE, 2013).

Research in corpus linguistics has led to the elaboration of better quality learner input and provided researchers and teachers with a wider, finer perspective into language in use (Campoy et al., 2010). Consequently, corpora are often used to select and prepare teaching materials (Walsh, 2010) and the use of corpus derived exercises have significantly enriched the learning environment (Aston, 1997). Moreover, McCarten (2007) states that the research in corpus linguistics has helped second language teachers to decide how many words, what kind of words and how many they need to teach their students. This means that the learning materials developed by the teachers can be made more specific, that is, by focusing on the most important words to be mastered by the students (cf. Liu & Jiang, 2015).

In relation to the selection of words to be used in teaching academic English, a number of teachers have turned to the Academic Word List (AWL) created by Coxhead (2000). In her research, Coxhead (2000) compiled 414 academic texts of more than 400 authors, comprising 3,513,330 tokens or running words and 70,377 types (individual words) in approximately 11,666 pages of text. To formulate this AWL, Coxhead filtered her word list with the words occurring in the GSL (General Service List). The GSL is the list of general words which are the most encountered words in any given general text (Nation, 2001). After filtering the words with the GSL, Coxhead finally concluded that there are 570 word families that play significant roles in academic texts. Moreover, Coxhead (2000) claims that "by highlighting the words that university students meet with a wide range of academic texts, the AWL shows learners with academic goals which words are most worth studying" (p. 213). Nevertheless, there are at least two problems with the AWL in representing the words used in academic texts. The first is the selection of the corpus. The corpus comprises four disciplines: arts, commerce, law, and science. These disciplines are not broad and inclusive enough. In addition, most of the texts were from New Zealand, i.e., sixty-four percent were sourced in New Zealand, 20% in Britain, 13% in the United States, 2% in Canada, and 1% in Australia (Coxhead, 2000).

Following the research conducted by Coxhead (2000), several other researchers have also tried to formulate other academic word lists. For instance, Browne et al. (2013) developed the New Academic Word List (NAWL). The NAWL consists of 963 headwords collected from an academic corpus of 288 million words. Different from Coxhead (2000) research, the data in the NAWL are from both oral and written corpora. Browne et al. (2013) compile the data from four sources. The first is the Cambridge English Corpus which comprises academic journals, non-fictions, student essays, and academic discourse available from Cambridge University Press. The second source is the Michigan Corpus of Academic Spoken English (MICASE). The third source is the British Academic Spoken English (BASE) corpus. The fourth source is the compilation of the 100 best-selling textbooks. However, the largest data for NAWL were from the Cambridge English Corpus (i.e., 86.3%), which could make the words included in the NAWL skewed towards a certain variety of academic texts.

Another example of an academic word list is the Academic Vocabulary List (AVL) was created by Gardner and Davies (2014). The AVL is derived from a 120-million-word academic sub-corpus of the 425-million-word Corpus of Contemporary American

English. So far, AVL is an academic word list built using the largest corpus. AVL is also a list that has the biggest number of words. There are 3000 lemmas listed in the AVL. A lemma is a set of lexical forms having the same stem and belonging to the same major word class, for the example: the lemma 'walk' consists of the words 'walk, walked, walking, and walks' (Baker, Hardie & McEnery, 2006). When we put these lemmas into word families, we found 1991 word families, which is too large to include in the teaching materials of English for academic purposes. We believe that learners need the smallest number of words, with the highest amount of coverage in the academic texts, especially in academic articles.

In spite of the criticisms, a number of learning materials have been produced using academic words derived from corpus linguistics research, notably from the AWL of Coxhead (2000). The *Oxford Learner's Dictionary of Academic English* (2014), for example, provides exercises that focus on the use of the AWL. For lecturers or teachers of English for Academic Purposes, the AWL has also been an important resource for creating teaching materials. Some examples for this can be found in the Academic Vocabulary website (Haywood, 2007). However, as teaching English for Academic Purposes covers various disciplines or subject areas, the vocabulary taught to the students may also be different. This means that teachers should also focus on the technical vocabulary, i.e., the words that are used in a particular discipline. The academic words commonly used in one subject area may also be different from those used in another subject area. Consequently, several authors have ventured into the creation of academic words in particular subject areas.

Some of the word lists that focus on particular disciplines are the Chemistry Academic Word List (Valipouri & Nassaji, 2013), the Nursing Academic Word List (Yang, 2015), and the Medical Academic Vocabulary List (Lei & Liu, 2016). Valipouri and Nassaji (2013) created the Chemistry Academic Word List based on a corpus of four million words which are taken from 1,185 chemistry research articles. Their study focused on a sub-discipline, i.e., Chemistry, and their use of the General World List (GSL) and the AWL as the stop lists actually created a technical vocabulary list for this particular sub-discipline. Yang (2015) collected 1,006,934 words taken from 252 full text articles of electronic nursing journals from the library of Chang Gung University of Science and Technology. This study also focused on a sub-discipline, i.e., nursing. The study carried out by Lei and Liu (2016), however, is more general, in the sense that it does not focus on a sub-discipline, but on a larger discipline, i.e., medical sciences. Lei and Liu (2016) compiled a 2.7 million-word corpus from medical journal articles (from Elsevier) and a 3.5 million-word corpus from medical English textbooks, so the total size of the corpus is 6.2 million running words.

The Demand for an Academic Article Word List

For the past several years, lecturers or teachers in several non-English speaking countries have been urged to publish in international journals that mostly use English. For the example in Indonesia, the Directorate General of Higher Education (Dikti) issued a Decision Letter number 152/E/T/2012 in 2012 that it is mandatory for all Indonesian postgraduate students to publish in international journals before they graduate (Dikti, 2016). Research in corpus linguistics can respond to this demand by creating a word list which is derived specifically from international journal articles. Most of the previous studies only focus on the general academic words, i.e., the words used across academic disciplines and academic texts (cf. Coxhead 2000). Several recent studies have focused on more specific disciplines, for example Lei and Liu (2016) for medical science, Valipouri and Nassaji (2013) for chemistry, and Yang (2015) for nursing. However, the

corpora are too specific, which means that the word list does not comprise of academic vocabulary, but technical vocabulary. In addition, the corpus is still a combination between journal articles and textbooks. Consequently, it is necessary to create a corpus from prominent international journal articles that can be used to formulate an Academic Articles Word List (AAWL). The word list will inform us of the words which are commonly used in international journal articles. English teachers will also know what words they need to teach to the students whose goal is to publish in English-language journals in the social sciences.

In creating an AAWL, we need to consider the classification of subject areas in international journals. One of the ways to identify the subject areas is by using a research database. In this study the research database chosen is Scopus, which is the largest abstract and citation database of peer-reviewed scientific journals, books and conference proceedings (Elsevier, 2017). Scopus publications are categorized into four subject areas: health sciences, life sciences, physical sciences, and social sciences. Given the differences in the coverage of each subject area, we can also assume that the academic words used in one subject area are different from those used in another subject area. Therefore, an AAWL should be bound to one particular subject area.

The selection of the subject area of the AAWL discussed in this article is based on the number of journals available in each subject area. In order to determine the number of journals, we can use the data available in one particular publisher which has the highest number of publications. A publisher that has the highest number of publications has thousands of journals, so a good number of articles can be obtained. In addition, by focusing on only one publisher, the classifications of the journals in the social sciences subject area will also be based on that particular publisher. Other publishers have their own criteria for classifying the journals and the subject areas. According to the data of the publishers indexed in Scopus (Elsevier, 2017), we can see that the publisher that has the highest number of publications is Elsevier, which is 10% of the total publications, followed by Springer (8%), Wiley-Blackwell (5%), Taylor & Francis (5%), and Sage (2%). Given these figures, the corpus of this study is derived from Elsevier journal articles. The information about journal articles published by Elsevier can be accessed from the ScienceDirect website.

Based on the data from the website (ScienceDirect, 2017), there are 3,845 journals titles available. From that number, 1,855 titles are under the subject area of health sciences, 1,221 titles for physical sciences, 1,165 titles for life sciences, and only 755 titles for social sciences. In this study, the subject area with the smallest number of titles is chosen, with the hope that the creation of the Academic Article Word List for Social Sciences (AAWL-SS) will assist non-English speakers to make publications and to have ideas on the research topics that they can write in the future. The AAWL-SS is expected to be a resource for international publications. Lecturers and researchers will be able to know the words or terms which are commonly used in international journal articles. Teachers will able to create teaching materials which are based on the current and more frequent words used in international journal articles. And, translators will also be helped in translating journal articles so that the choice of words, among various synonyms, will be more suitable for the particular subject area.

The Formulation of the Academic Article Word List for Social Sciences

The corpus of this study is derived from journal articles in the social sciences that meet the following two criteria. The first criterion is that the articles are from the journals that provide open access articles. This is necessary to comply with the copyrights issues because non-open access articles cannot be freely copied and used without written

permission from the publisher. In Elsevier, there were 237 journal titles in social sciences that have open access articles. The second criterion is that the articles selected must be from a journal that has a 5-year impact factor. This is the sign that the journals have been available and cited for a good number of years.

From the 237 journal titles in social sciences that have open access articles, some of them are also categorized under other subject areas. Since our focus is on the social sciences, we excluded the journal titles that are categorized across several subject areas, i.e., not only categorized in social sciences. Appearing across several subject areas mean that the journals are not purely social sciences. One example of such a journal is the one entitled *Alcoholism and Drug Addiction* which is categorized in both health sciences and social sciences. After this selection process, we found 122 journal titles that are so-called purely within the subject area of social sciences. From each of these journals, we selected only one article, so that every journal was presented equally in the corpus. The journal articles selected were those published in 2012-2016 to obtain the recent publications. These journal articles were in a PDF format and had to be converted into a .txt format for further processing. The software called AntFileConverter (Anthony, 2015) was used to convert these PDF files into txt files.

As with any results of converting software, there are always some typos that need to be corrected after the converting process. In this case, we read all the converted articles and compared them with the original documents in order to make corrections for the words or letters which were converted incorrectly. The final result was a corpus of 1,040,259 word tokens.

After obtaining the corpus, we created the word list using the software called AntWordProfiler (Anthony, 2014). The word list was created by considering three conditions. For the first condition, we used the GSL as a stop list. A stop list means the list of words which are excluded from the calculation. This allowed us to exclude the general words automatically. This same technique was also used in the creation of the AWL (Coxhead, 2000). For the second condition, we only took the words with the range of minimum six. This is due to the fact that the subject area of social sciences is further categorized into six sub-disciplines in the Elsevier journals, they are (1) Arts and Humanities, (2) Business, Management and Accounting, (3) Decision Sciences, (4) Economics, Econometrics and Finance, (5) Psychology, and (6) other Social Sciences. We would like the words to exist in all of these six sub-disciplines, so that the collected words represent those used in the whole sub-disciplines of social sciences. This means that all of the sub-disciplines of social sciences will be equally represented in the corpus. For the third condition, after complying the previous two conditions, the final words were selected based on their frequencies.

We finally decided that there are 350 words included in the Academic Article Word List for Social Sciences. These 350 words are in the form of word families, following the method used in the AWL. To make the word members (i.e., the inflections and derivations) of these words, we used the software called Familizer. This software can be accessed for free from the Lextutor website (Cobb, 2016). Figure 1 shows how the web page appears when the words have been added into the website.
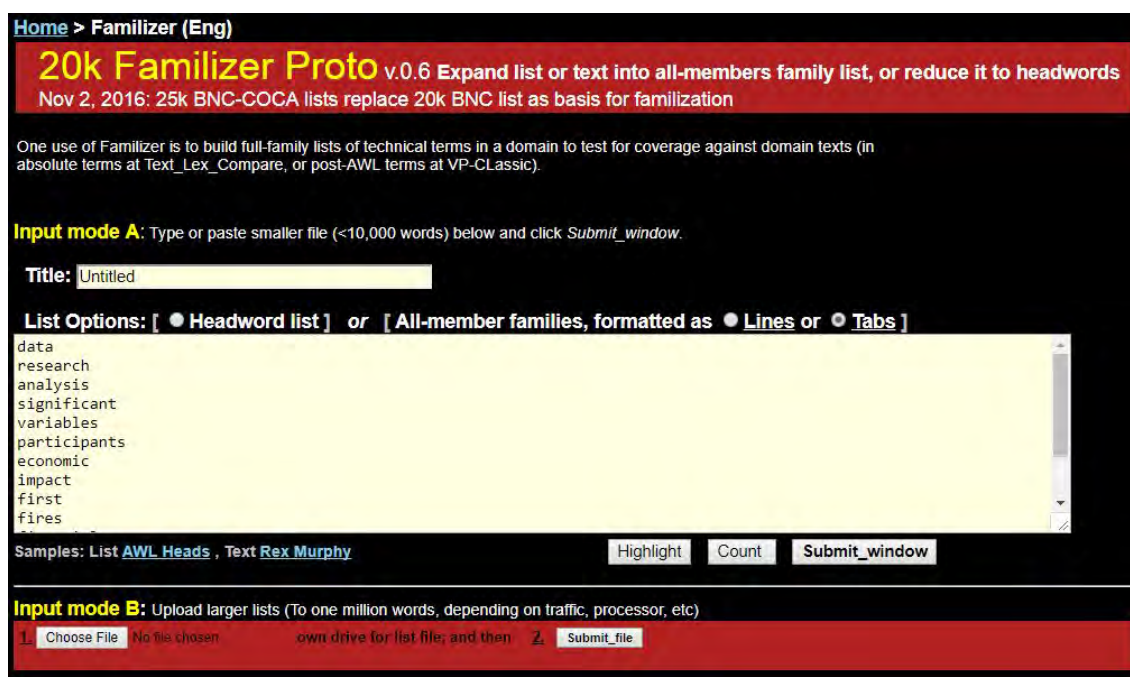
Figure 1. Familizer web page added with the data.

The result of the programming using Familizer is a text file with the word members in tab format (see Figure 2). This means that every word family will be shown with all its inflections and derivations. For example, the word family 'converge' will be shown with its inflections (i.e., converges, converged, and converging) and its derivations (i.e., convergence and convergent).
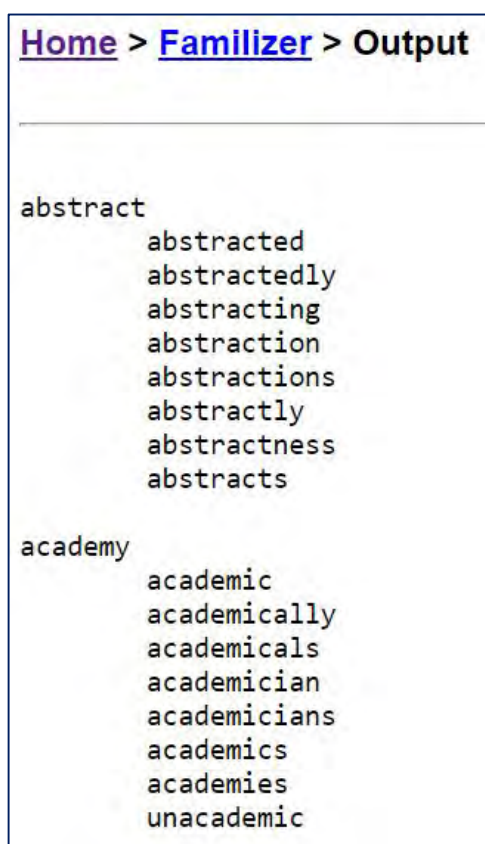

Figure 2. An extract of the list created using the Familizer.

This file produced by the Familizer can be used as a stop list in when corpus linguists want to test the coverage of the words in a text. This means that corpus linguists can use this stop list file in the software such as Range (Nation, 2005) or AntWordProfiler

(Anthony, 2014). We have provided this stop list file to be freely downloaded from the AAWL-SS website (Kwary, 2017).

Testing the Academic Article Word List for Social Sciences

After creating the AAWL-SS, the next step is to test the coverage of the words listed in the AAWL-SS. In testing this word list, we compare the coverage of this AAWL-SS with three other word lists, i.e., the AWL as the representative of the classic academic word list, the NAWL as the newer version of the academic word list, and the BNC-COCA baselist3 (British National Corpus – Corpus of Contemporary American English). The AWL is already in the form of word families, so it can be used directly. The NAWL is in the form of headwords (only the words with their inflections), so we need to convert the NAWL headwords (i.e., 963 headwords) into word families (containing not only the inflections, but also the derivations). For example, the words 'accumulate' and 'accumulation' in the NAWL are counted as two lemmas, but when they are converted into word families, they belong to only one word family, i.e., accumulate. In this case, we use the software called Familizer to make the word families, and the result is 880 word families. The number of words is smaller because some headwords can be joined together to form a word family.

The BNC-COCA baselist3 (which will be called BNC-COCA3 in this article) is used to determine whether the AAWL-SS contains specific vocabulary, or simply general vocabulary. The BNC-COCA contains a number of baselists, which are created from two big general corpora, i.e., the British National Corpus and the Corpus of Contemporary American English. The list was created by Paul Nation based on the frequency of the words in those two corpora. The complete list can be downloaded from his personal website (Nation, 2016). The first two baselists, called baselist1 and baselist2, are not used in this study because they are parallel with the GSL, which is used as a stop list in creating the AAWL-SS.

The first test run to determine the coverage of the AAWL-SS used the corpus which was previously built, i.e., the 1,040,259 words, taken from 122 journal articles within the subject area of social sciences. The calculation was done automatically by using AntWordProfiler (Anthony, 2014). The recapitulation of the coverage of the AWL and that of AAWL-SS is presented in Table 1.

File	Token	Token %
AWL	122,296	11.76
NAWL	108,031	10.39
BNC-COCA3	107,361	10.32
AAWL-SS	131,574	12.65

Table 1. The coverage of AWL, NAWL, BNC-COCA3 and AAWL-SS in the 1,040,259 word tokens.

As shown in Table 1, the coverage of the AAWL-SS is higher than those of the AWL, NAWL, and BNC-COCA3. There are 131,574 word tokens, from the 1,040,259 word corpus, which are included in the AAWL-SS, and there are only 122,296, 108,031, and 107,361 word tokens which are covered by the AWL, NAWL, and BNC-COCA3, respectively. The difference in the percentage looks small, i.e., only about 1% against the AWL (i.e., 12.65% for the AAWL-SS and 11.75% for the AWL). However, we need to note that the number of word families in the AWL (570 words) is higher than that of AAWL-SS (350 words). The difference is more than 2% against the other two lists, i.e., the NAWL and the BNC-COCA3. The 2% difference may look small, but it is 2% from the 1,040,259 word corpus, so in term of number of words, the difference is

approximately 20,000 words. This means that the 350 word families in the AAWL-SS have a better coverage (by thousands of words) of the academic words used in journal articles of the social sciences than the AWL, NAWL, and BNC-COCA3. In other words, with the smaller number of words, i.e., only 350 word families, the learners can actually obtain better coverage of the words used in journal articles, i.e., 12.65% or 131,574 word tokens.

Several other researchers (cf. Browne et al., 2013; Gardner & Davies, 2014) have also used the same technique of comparing the formulated word list with the initially collected corpus. The results are similar to that shown in this study, i.e., the formulated word list provides better coverage of the text. However, one may argue the use of the initially collected corpus in determining the coverage of the words. The word list was derived from the initially collected corpus, so the coverage must also be high in that particular corpus. To answer this argument, we made another calculation using a different corpus to confirm the coverage of the AAWL-SS. In this second test, we collected different 25 journal articles of the social sciences from the same publisher. This new corpus contains 252,114 word tokens. The recapitulation of the coverage of the AWL, NAWL, BNC-COCA3, and AAWL-SS is shown in Table 2.

File	Token	Token %
AWL	30,342	12.04
NAWL	26,173	10.38
BNC-COCA3	25,571	10.14
AAWL-SS	32,681	12.96

Table 2. The Coverage of AWL, NAWL, BNC-COCA3 and AAWL-SS in the 252,114 word tokens.

The result presented in Table 2 confirms the previous conclusion that the AAWL-SS has a better coverage than the other word lists, in spite of the fact that the AAWL-SS consists of the smallest number of word families, i.e., only 350 word families. In Table 2, we can see that the AAWL-SS covers 32,681 word tokens in this second corpus, while the AWL only covers 30,342 words, there are 26,173 words for the NAWL, and there are 25,571 words for the BNC-COCA3. The difference in the percentage is again small, i.e., 12.96% for the AAWL-SS and 12.04% for the AWL, and a bit higher for the AAWL-SS against NAWL and BNC-COCA3. However, as stated previously, the difference in the number of word families is quite high, so it shows the significance of the AAWl-SS over the AWL, NAWL, and BNC-COCA3. Naturally, learners would prefer learning 350 words to learning 570, 880, or 1,000 words if both result in the same coverage. The superiority of the AAWL-SS over the other word lists in this case can be due to the fact that the AAWL-SS was specially designed for journal articles in the social sciences. The AWL covers more general academic words. The NAWL covers an even more general and broader academic genre, and the BNC-COCA3 covers general words.

If we look closely at the word families listed in AWL and AAWL-SS, we can indeed find some differences. Our calculation shows that there are 262 word families that can be found in both the AWL and the AAWL-SS. Some of these words are *analyse, create, distribute, establish,* and *indicate*. There are 308 word families found in the AWL but not in the AAWL-SS. Some of the examples are *acknowledge, bond, interval, ministry*, and *tape*. Finally, there are 88 word families that can be found in the AAWL-SS, but not in the AWL. This includes the words *cognition, converge, dense*, *robust* and *volatile*. Consequently, the AAWL-SS is a better option for teachers who want to create materials for teaching students of social sciences about writing articles for international journals.

In addition to the words which are formulated based on their word families, we also created a list of words based on their word types (see Appendix 1). The word types placed in the appendix are the word types with the highest frequency. For example, in the word family members, the word *converge* is listed with the following words: *converges, converged, converging, convergence,* and *convergent*. However, for the word type, our calculation on the frequencies shows that the word that has the highest frequency for this word family is *convergence*. Therefore, in the list shown in the appendix, only the word *convergence* is included. Another example is the word family of *culture* which consists of *cultural, culturally, cultured, cultures,* and *uncultured*. The calculation of the frequencies of the word types shows that the adjective *cultural* is more frequently used than the other word types. This means that creating a vocabulary exercise using the word *cultural* will be more useful for the students than using the other word types. In other words, the list of word types will enable teachers to make even more specific teaching materials, particularly vocabulary exercises for students from social sciences who want to make publications in international journals.

Conclusion

This study has shown that the newly created Academic Article Word List for Social Sciences has a better coverage of the vocabulary used in social sciences journal articles than the Academic Word List, the New Academic Word List, and the BNC-COCA baselist3. This can be due to the fact that the AAWL-SS is more specific to one subject area, i.e., social sciences, and to one type of academic text, i.e., international journal articles. This more focused approach results in a smaller number of word list, i.e., 350 word families, but with a better coverage of the vocabulary. Consequently, the AAWL-SS is a valuable reference for teachers, textbook authors, translators, and researchers who want to focus on the words most frequently used in social sciences journal articles. Teachers and textbook authors will be able to create teaching materials which are more focused on the words needed by students who hope to publish in English-language journals in the social sciences. Translators will be able to select the right words among various synonyms, which are more commonly used in social sciences journal articles. And, researchers will be able to explore the words and the behaviour of the words commonly used in journal articles, particularly in the field of social sciences.

References

Aijmer, K., & Altenberg, B. (1991). *English corpus linguistics: Studies in honour of Jan Svartvik*. London, UK: Longman.

Alsop, S., & Nesi, H. (2009). *Issues in the development of the British Academic Written English (BAWE) corpus.* Corpora, 4(1), 71-83. doi: 10.3366/E1749503209000227

Anthony, L. (2014). *AntWordProfile*r (Version 1.4.1) [Software]. Available from http://www.laurenceanthony.net

Anthony, L. (2015). *AntFileConverter* (Version 1.2.0) [Software]. Available from http://www.laurenceanthony.net

Aston, G., (1997). Enriching the Learning Environment: Corpora in ELT. In A. Wichmann, S, Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 51-64). London, UK: Longman.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh, Scotland: Edinburgh University Press.

Bogaards, P., & Laufer, B. (2004). *Vocabulary in a second language*. Amsterdam, Netherlands: John Benjamins.

Browne, C., Culligan, B., & Phillips, J. (2013). *A new academic word list*. Available from http://www.newgeneralservicelist.org/nawl-new-academic-word-list

Campoy, M.C., Belles-Fortuno, B. and Gea-Valor, M.L. (Eds.), (2010). *Corpus-based approaches to English language teaching*. London, UK: Continuum.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.). New York, N.Y.: Routledge.

Cobb, T. (2016). *20 K Familizer Proto* (Version 0.6) [Software]. Available from http://www.lextutor.ca/familizer

Coxhead, A. (2000). *A new academic word list*. TESOL Quarterly, 34(2), 213-238. doi: 10.2307/3587951

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, 25(4), 447-464. doi: 10.1093/llc/fqq018

Dikti (2016). *Pangkalan Data Pendidikan Tinggi*. Available from https://forlap.ristekdikti.go.id

Elsevier. (2017). *Scopus*. Available from https://www.elsevier.com/solutions/scopus

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327. doi: 10.1093/applin/amt015

Haywood, S. (2017). *Academic vocabulary*. Available from https://www.nottingham.ac.uk/alzsh3/acvocab/teaching.htm

Kwary, D. (2017). *AAWL-SS*. Available from http://corpus.kwary.net/freq

Laviosa, S. (1998). The corpus-based approach: A new paradigm in translation studies. *Meta: Translators' Journal*, 43(4), 474-479. doi: 10.7202/003424ar

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53. doi: 10.1016/j.jeap.2016.01.008

Liu, D., & Jiang, P. (2015). Corpus-Based lexicogrammatical approach to grammar instruction: Its use and effects in EFL and ESL contexts. In M. Christison, D. Christian, P.A. Duff, & N. Spada (Eds.), *Teaching and learning English grammar: Research findings and future directions* (pp. 103-118). New York, N.Y.: Routledge.

McCarten, J. (2007). *Teaching vocabulary: Lessons from the corpus, lessons for the classroom*. Cambridge, UK: Cambridge University Press.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, I. S. P. (2005). *Range program with GSL/AWL list*. Available from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Range_GSL_AWL.zip

Nation, I. S. P. (2016). *Range program with BNC/COCA lists 25,000 words*. Available from http://www.victoria.ac.nz/lals/about/staff/publications/BNC_COCA_25000.zip

ScienceDirect. (2017). *ScienceDirect*. www.sciencedirect.com/science/journals

Sinclair, J. (2005). Corpus and text: Basic principles. In Wynne, M. (ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxford, UK: Oxbow Books. http://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, 12(4), 248-263. doi: 10.1016/j.jeap.2013.07.001

VOICE. (2013). *The Vienna-Oxford international corpus of English*. http://voice.univie.ac.at

Walsh, S. (2010). What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses? In A. O'Keeffe & M. McCarthy (Ed.) *The Routledge handbook of corpus linguistics* (pp. 333-344). New York, N.Y.: Routledge.

Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, 37, 27-38. doi: 10.1016/j.esp.2014.05.003

Appendix 1.

The 350 Word Types of the Academic Article Word List for Social Sciences

abstract, academic, access, accuracy, achieve, acquisition, additionally, adjusted, adults, affect, aggregate, allocation, alternative, analysis, annual, appendix, approach, appropriate, approximately, areas, aspects, assess, assets, assigned, assumption, attitudes, attribute, authority, available, awareness, baseline, bias, brand, capacity, capture, cash, category, causal, challenges, classroom, climate, clusters, coded, cognitive, column, communication, community, competence, competitive, complex, component, comprehension, computational, concept, conclusion, conducted, consensus, consequences, consistent, constant, constraints, construct, consumption, contemporary, context, contrast, contribution, convergence, cooperation, core, corporate, correlation, corresponding, creation, credit, crisis, criteria, crucial, cultural, cycle, data, dataset, decline, demographic, demonstrate, density, dependence, derived, design, despite, destination, determinants, developmental, deviation, diagnostic, dialogue, digital, dimensions, discourse, distinction, distribution, diversity, documented, domain, domestic, dominant, dummy, duration, dynamic, economic, elements, emerging, emotional, emphasis, empirical, enable, energy, engage, enhanced, ensure, environment, equation, equilibrium, equity, errors, established, estimates, evaluation, evidence, evolution, expenditure, explicitly, exposure, external, facilities, factors, features, feedback, finally, financial, focus, framework, function, funding, furthermore, gender, generate, geography, global, goal, goods, governance, grade, guidelines, hazard, hence, highlight, household, hypothesis, identify, illustrate, impact, implementation, implications, implies, incentive, income, index, indicate, individual, infants, inference, infrastructure, initial, innovation, insights, instance, institutions, instruction, integration, intelligence, intensive, interaction, internal, interpretation, intervention, interviews, investigate, investment, involved, issue, items, job, labour, leverage, linear, linguistic, link, literacy, location, maintenance, major, marginal, matrix, maximum, mechanisms, media, median, mental, method, migrants, minimum, mobility, monitoring, narrative, negative, network, neutral, nevertheless, normative, novel, objective, obtained, occupation, occur, online, optimal, option, organic, organizational, outcomes, output, overall, panel, parameters, participants, partner, payoff, peer, percentage, perceptions, period, perspective, phase, physical, platforms, policy, portfolio, positive, potential, predicted, previous, primary, prior, procedure, process, professional, project, promote, proportion, proxy, published, purchase, qualitative, quantitative, questionnaire, random, range, ratio, recall, recipient, reference, regional, regression, relevant, required, research, residential, resources, respondents, restricted, retail, revealed, robustness, role, rural, scenario, scheme, scholars, score, section, sector, security, seeking, selected, series, settings, shift, signaling, significant, similar, sites, software, source, spatial, specific, stability, statistically, status, strategies, structure, subjective, subsequent, summary, survey, sustainable, symptoms, systemic, target, task, team, technical, techniques, technology, text, theory, threshold, tourism, traditional, transactions, transfer, transitions, transport, underlying, unique, urban, utility, validity, variables, verbal, version, versus, via, video, visual, vocabulary, volatility, web, whereas, zone.