

An English Cover Letter Essential Wordlist for Second Language Learners

MacPaul Hirata

Kanda University of International Studies

Writing a cover letter is an essential part of the job application process. To find ways to improve second language learners' cover letter writing ability, this author examined cover letters to create a list of words that are essential for writing English cover letters. A Cover Letter Essential Word List (CLEWL) of 347 words was created from a corpus of 400 cover letters. The CLEWL was analyzed in terms of its make-up, lexical coverage, and lexical frequency profile. This study presents those findings, as well as suggestions for teaching words from the CLEWL.

カバーレターを書くことは就職活動に不可欠な部分である。本論では第2言語学習者のカバーレター作成能力を向上させる方法を見つけるために、英語のカバーレターの文例を調査し、カバーレターを書く上で不可欠な単語リストを作成した。400枚のカバーレターのコーパスから、カバーレターの必須単語リスト (CLEWL) 347語を作成し、それを構成、語彙の範囲、および語彙頻度プロファイルの観点から分析した。本論ではその調査結果について述べ、CLEWLの単語を教える方法を提案する。

The purpose of this study was to determine which words would be most useful for L2 learners to know when they are composing a professional cover letter. The author looked at cover letters from a corpus linguistics perspective with the goal of creating a specialized word list for this text type. A cover letter is a document sent with applicants' resumes to a potential employer to provide additional information about their qualifications. Cover letters transverse every discipline and remain widely used. The significance of this study is that there is, to date, no study on the essential vocabulary for cover letters despite the benefits of such an endeavor. Words lists such as this one are important for helping L2 learners to gain both specialized receptive vocabulary knowledge and productive vocabulary knowledge (Yamamoto, 2014).

Making an Area-Specific Word List

With the availability of computing resources and the subsequent advancement in computer technology, computer programs have become the tool of choice for making specialized word lists. This is because of their ability to process a large amount

of data in a limited time and their consistency in objectively identifying technical and non-technical terms when compared to other methods, such as compilation by human researchers and the use of technical dictionaries. When using computer programs to develop specialized word lists, four criteria can be applied. These are *range*, *frequency*, *dispersion*, and *ratio*.

Range

Range is the number of different texts in which a word appears. Both Coxhead (2000) and Nation (2016) rank range as the most important criteria. The reason is that range avoids the bias of the frequency criteria where a word might appear many times within a text, probably because of its relationship to the topic of the text, and yet fail to appear in other different texts in the same corpus. The more useful a word is, the more likely it is to appear in many different texts. Coxhead (2000), when making the Academic Word List (AWL), set the threshold of occurrence to 15 or more of the 28 texts in the corpus, while Yang's (2015) threshold for the inclusion of words in the Nursing Academic Word List (NAWL) was occurrence in 11 of the 21 subject areas of the Nursing Research Articles Corpus (NRAC).

Frequency

Frequency refers to the number of times a word occurs within each text. However, because topic-related words could have a high frequency in a single text, and yet have a limited range, frequency alone might not be sufficient criteria for the inclusion of a word. The generally accepted frequency threshold for a word is 28.57 times per million words, or 0.0028% of the total running words in a corpus (Coxhead, 2000; Lei & Liu, 2016; Yang, 2015).

Dispersion

The *dispersion* criterion is used to eliminate the bias of the frequency count by making sure that words occur evenly across the different texts in a corpus. The dispersion of a word in a corpus is acquired by

dividing its range and frequency data. Gardner and Davies (2014) used a dispersion threshold of 0.80 in creating their Academic Vocabulary List.

Ratio

The *ratio* criterion compares the frequency rate of the selected words in a target corpus with their frequency in another non-related corpus. The reason for this is the belief that items in a specialized list are more likely to occur more frequently in related texts than in other non-related texts. Gardner and Davies (2014) adopted a threshold of a 1.5:1 ratio in their Academic Vocabulary List. This means that the selected words had to appear in the specialized text at least 50% more than in non-related text.

The above four criteria are used in varying combinations in different studies on specialized word lists. For instance, Coxhead and Hirsch (2007) used a combination of range, frequency, and dispersion to determine the technical vocabulary of science texts. Wang, Liang, and Ge's (2008) Medical Academic Word List (MAWL) and Yang's (2015) NAWL were both based on range and frequency. Frequency, dispersion, and ratio were used by Zhu (2017) in his study of the technical vocabulary of newspapers, and by Lei and Liu (2016) to determine the technical vocabulary of medical texts.

In the present study, range, frequency, and dispersion were used to select words that are essential for writing cover letters. Ratio was not calculated, because this study was not concerned with a comparison of the cover letter corpus with another non-related corpora.

Research Questions

Two research questions that form the basis of this study are:

1. Is there a list of words that are essential for writing English cover letters?
2. What is the lexical frequency profile of the Cover Letter Essential Word List?

Methods

Adopting the computer-based approach (Chung & Nation, 2004), a list of frequently occurring words was generated from a corpus of cover letters (CL). The result formed the Cover Letter Essential Word List (CLEWL). Because cover letters are quite personal documents, unlike other types of text such as newspaper articles, which are made for public consumption, this study relied mostly on online sources for the cover letter samples. These online sources

were educational websites that offered training and advice on job applications. Three websites, *dayjob.com*, *bestsamplersresume.com*, and *monster.com*, were included. Thus, while the samples might not have constituted authentic cover letters, they were no less representative of the actual form of this text type. The corpus consisted of 406 CL samples with a total of 83,196 running words. The samples varied in size from 300 to 600 words and were taken from 11 different fields of work (see Table 1) to ensure diverse representation. However, to make the samples computer readable, the cover letters from each field were merged into one large file.

Table 1. List of Job Fields and Their Corresponding Number of CLs

Field	Number of CLs	Tokens
Aviation	36	4,471
Education	37	7,265
Engineering	34	6,062
Finance	44	8,891
Health	40	9,930
Human resources	38	8,879
IT	36	7,101
Marketing & Advertising	34	7,874
Mass Media	37	8,425
Real Estate	36	6,640
Hospitality	34	7,658

At the processing stage, the samples were standardized by removing unnecessary information, such as addresses, and dates, which were of no value to the lexical analysis of the samples. Hyphenated words that had a different meaning from those of their individual words were moved into an off-list. Also, using a stop list, function words and proper nouns were eliminated from the samples because their meanings were independent of the subject matter. The WordSmith Tools (Version 7.0; Scott, 2018) program was used to determine range, frequency, and dispersion among the selection of words, and the intuition and knowledge of four human validators was relied upon to classify the make-up of the CLEWL.

The Range computer program (Heatley, Nation, & Coxhead, 2002) was used to obtain the lexical frequency profile of the resulting CLEWL. The lexical frequency profile shows how many items of

the CLEWL are distributed at different 1,000-word lexical frequency bands. This involved running the CLEWL as a text in the Range program, first against the GSL-AWL list, then with the BNC-COCA list. The GSL-AWL is a combination of West's (1953) list of 2,000 general high frequency English words, and Coxhead's (2000) 570-word family list of English academic vocabulary. The BNC-COCA (Nation, 2012) is a list of the 25 most frequent 1,000 word families common in English text.

Finally, according to Nation (2016), when developing a word list, "the unit of counting needs to reflect the kind of word knowledge needed by the end-users of the list" (p. 8). Because the end-users of the CLEWL are L2 learners who are writing cover letters, word type was used as the unit of counting in this study. A word type refers to every different word in a text or corpus. For instance, *write* and *writing* are counted as two different words. In comparison to tokens and word families, word types are more suitable for making word lists for productive knowledge (e.g., speaking and writing), because of the necessity for knowledge of the individual word forms and their appropriate use contexts (Nation, 2016).

Results

The CLEWL that was created from the CL corpus consisted of a total of 347 word types. These occurred 15,930 times and made up 29.31% of the total running words in the corpus. The most frequent 20 words of the list alone made up 4,269 tokens or 5.15% of the total running words in the corpus (see Table 2).

Also, 77.52% of the CLEWL words were contained in the GSL-AWL, and 89.91% were contained within the top 10 frequency bands of the BNC-COCA word list. This means that the CLEWL is composed mostly of high-to-mid frequency words.

Discussion

Henry and Roseberry (2001) note the importance of teaching learners genre-specific vocabulary because it will help to determine conformity to the conventions of the genre. This is particularly necessary in the cover letter subgenre because the success of the writer depends on adherence to the rules of the genre. Consequently, this study found a list of 347 word types that were determined to be essential for the writing of English cover letters. The CL-essential vocabulary was found to account for 29.31% of the total running words in the CL corpus. However, the less specialized nature of the CL subgenre means that its vocabulary would be more generally useful than, for example, vocabulary from a science-related

Table 2. Most Frequent 20 Words of the CLEWL

Word	Frequency	%
Experience	464	0.56
Position	395	0.47
Resume	339	0.41
Skills	279	0.34
company	257	0.31
Years	255	0.31
Job	223	0.27
Time	213	0.26
Organize	196	0.24
Interview	187	0.22
Service	168	0.20
Contact	163	0.20
Meet	163	0.20
Qualifications	161	0.19
Team	161	0.19
Enclosure	157	0.19
Application	125	0.15
Forward	124	0.15
Look	121	0.15
Discuss	118	0.14

Table 3. Make-Up of the CLEWL

Level	Description	Examples
1	Words that have no specific relationship with CLs	function words, proper nouns
2	Words that have meanings that are minimally related to CLs.	<i>high, highly, good, excellent, increase, last, previous</i>
3	Words that have meanings that are closely related to CLs.	<i>experience, interests, chance, motivated, goals, ability</i>
4	Words that have meanings that are specific to CLs	<i>CV, resume, skills, position, interview, advertised, job</i>

discipline. That is, the words can have the same exact meaning and usage in other non-CL texts.

The Make-Up of the CLEWL

The CLEWL was evaluated using Chung and Nation’s (2003) four-point rating scale (see Table 3).

None of the words in the list belonged to the first level because the stop list of function words and proper nouns that was created at the data processing phase eliminated these words. Few words could be classified under the second level, and these words accounted for less than 5% of the words in the list. Most of the words from the list can be classified under level 3. These are words such as *motivate*, *chance*, and *achieve*, which are essential for writing effective CLs, but which are not specific to the subgenre. On the other hand, far fewer words were classified as level 4 words compared to level 3 words, and these were words that were considered specific to CLs. However, these are also likely to be found with the same meaning in other fields. The concentration of the essential words in level 3 attests to the less specialized nature of the vocabulary of CLs in general.

To validate the above classification, two English for specific purposes (ESP) instructors and two general English instructors rated the list and agreed that 70–80% of the CLEWL could be rated under level 3.

CLEWL Lexical Frequency Profile

An analysis showed that 77.52% and 89.91% of the CLEWL are contained in the GSL-AWL and within the first 10 frequency bands of the BNC-COCA. Less than three percent can be found at the low-frequency level of the BNC-COCA word list. This

means that the CLEWL is made up mostly of high to mid-frequency level words. Because beginner and pre-intermediate level learners are unlikely to know a considerable number of these words, the CLEWL could be an effective learning resource for ESP students learning how to write English cover letters. Table 4 shows some of the CLEWL words within the GSL-AWL lists. Learning the essential vocabulary of cover letters is important because it can help to reduce learners’ lexical burden and make it easier for learners to write effective cover letters.

Table 4. CLEWL Items in the GSL-AWL

1k	2k	AWL
experienced	advertisement	appropriate
applying	confident	challenges
opportunities	enclosed	dynamic
expectations	managing	energetic
current	improving	motivated
writing	education	role
training	match	professional
efficient	perform	schedule
fit	discuss	unique
lead	skills	tasks
position	qualifications	strategies

Teaching Implications of the CLEWL

One way in which the CLEWL could be taught to ESP learners would be to organize the words into the various parts of speech. For example, the words could be grouped under action verbs, self-descrip-

Table 5. Cover Letter Functions and Corresponding Vocabulary

Offering Candidature	Self-Promotion	Enclosing Documents	Welcoming Response
write	developed	attached	contact
apply	skills	CV	interview
position	possess	Resume	look forward
contact	efficient	proof	appreciate
interest	planned	enclosed	opportunity
advertized	managed	references	chance
express	experienced	information	consideration
	education	include	response
	current		convenience

tive adjectives, abstract nouns, and adverbs. Another way would be to group the words according to the functions that they serve in a cover letter (see Table 5).

There is much value in collecting and organizing the essential vocabulary of cover letters, as this study has done, before learners start to learn cover letter writing. If they can learn these words in advance, learners will then be able to focus more on learning the technical vocabulary of their field or industry of interest.

The main limitation of this study is the nature of the text type. CLs are personal documents and are usually available only to the writer and the target recipient. This makes access to authentic CLs very challenging. The data used in this study were acquired mostly from contrived samples which might have influenced the make-up of the CLEWL. Another limitation stems from the problem of socio-cultural differences among countries. These differences may cause variations in CL writing style from one country to another. For example, British CLs might differ from American CLs. Most of the sample CLs in this study were taken from American websites. A more representative corpus must contain samples of CLs from different countries. Also, while the 400 cover letters that make up the CL corpus might seem large, a much larger sample would insure a high representation of low-frequency words.

Conclusion

This study investigated two important research questions. The first was whether there is a set of words that are essential for writing cover letters. A cover letter essential word list (CLEWL) of 347 word types was extracted from a corpus of 400 cover letters based on range, frequency, and dispersion. Data showed that the CLEWL accounted for 29.31% of the total running words in the corpus. The second question examined the lexical frequency profile of the CLEWL by comparing it with the GSL-AWL and the BNC-COCA lists. The CLEWL was found to cover high- to mid-frequency band words and was therefore a useful resource for students in learning how to write cover letters. Given the importance of cover letters in every professional field, more studies on the essential vocabulary of cover letters using a larger corpus are needed.

References

- BSR. (n.d.). *Cover Letters*. Retrieved from <https://www.bestsamplesresume.com/resume-cover-letter.html>
- Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15(2), 103–116. Retrieved from <http://nflrc.hawaii.edu/rfl/October2003/chung/chung.html>
- Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32(2), 251–263. doi:10.1016/j.system.2003.11.008
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. doi:10.2307/3587951
- Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, 12(2), 65–78. Retrieved from <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2007-2-page-65.htm>
- DayJob.com. (n.d.). *Cover letter examples*. Retrieved from <https://www.dayjob.com/content/cover-letter-examples-314.htm>
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. doi:10.1093/applin/amt015
- Heatley, A., Nation, I. P., & Coxhead, A. (2002). Range [Computer software]. Retrieved from http://www.vuw.ac.nz/lals/staff/Paul_Nation
- Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: “Letter of application.” *English for Specific Purposes*, 20(2), 153–167. Retrieved from <https://www.sciencedirect.com/science/article/pii/S088949069900037X>
- 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
- Monster.com. (n.d.). *Cover letters and resumes*. Retrieved from https://www.monster.com/career-advice/cover-letter-resume?intcid=skr_navigation_www_cover-letter-resume
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Philadelphia, PA: John Benjamins Publishing Company.
- Scott, M. (2018). *WordSmith Tools* [Computer software]. Oxford, England: Oxford University Press.
- Wang, J., Liang, S.-L., & Ge, G.-C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442–458. doi:10.1016/j.esp.2008.05.003
- Waring, R., & Nation, I. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 6–19). Cambridge, England: Cambridge University Press.
- West, M. (1953). *A general service list of English words*. London: Longman, Green, & Co.

Yamamoto, Y. (2014). Multidimensional vocabulary acquisition through deliberate vocabulary list learning. *System*, 42(1), 232–243. doi:10.1016/j.system.2013.12.005

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

Zhu, J. (2017). *The technical vocabulary of newspapers* (Master's thesis, University of Western Ontario, Ontario, Canada). Retrieved from <https://ir.lib.uwo.ca/etd/4872>

MacPaul Hirata is a lecturer in the ELCC department of Kanda University of International Studies. He holds an M.A. in Applied Linguistics and TESOL from the University of Leicester, UK. His primary research interest is L2 vocabulary teaching and learning. He is also interested in technology integration in language learning. He can be contacted at hirata-m@kanda.kuis.ac.jp.



Appendix

The appendix of the complete *Cover Letter Essential Word List* can be found in the online version of this article at <http://jalt-publications.org/tlt/>.

NEAR

北東アジア言語教育学会
Language Education Conference

NEAR Conference

June 1, 2019, Niigata

Join us in Niigata for the 11th annual NEAR conference, the premiere language education event on the Japan Sea coast. The call for papers is open now! <http://nearconference.weebly.com/>

As we enter our 2nd decade, we are proud to announce our new peer-reviewed conference proceedings.

Come to Niigata! Present your work! Publish your results!

JALT2019 Volunteering JALT2019 Reading Committee

“Wanna help with planning the 2019 international conference and educational materials exhibition in Nagoya? Of course you do!! Well, contact me, Wayne Malcolm, new Director of Program at wamalcolmjalt2008@gmail.com and I’m sure we can find something for your to do. In fact, we are currently looking for individuals interested in joining the JALT2019 Reading Committee, responsible for vetting presentation proposals.

This is a great way to get engaged with JALT.

We want you!! Hope to hear from you soon!”



Never had an article published before? Lacking confidence, or just unsure of what to do?

TLT’s Peer Support Group can help.

<jalt-publications.org/psg/>

Learn to write with the help of our experienced collaborative writing team.