Initial Evaluation of AWSuM: A Pilot Study

Atsushi Mizumoto

Kansai University

doi: <http://dx.doi.org/10.7820/vli.v06.2.Mizumoto>

Abstract

This article reports on a pilot study to evaluate usability, effectiveness, and impact of Academic Word Suggestion Machine (AWSuM), a web-based, innovative tool, powered by the combination of rhetorical moves and lexical bundles with an auto-complete feature. Eight L2 writers participated in the study. Their open-ended comments after using the tool were qualitatively analyzed and classified in thematic categories. The results showed that the developed tool brought about beneficial effects that genre writing pedagogy explicitly aims to achieve. Implications are discussed in terms of the potential role that a tool like AWSuM could play in the teaching and learning of technology-enhanced genre writing.

1 Academic Word Suggestion Machine

Academic Word Suggestion Machine (AWSuM) is an online tool that provides support for English academic research writers through an approach that integrates move analysis (Swales, 1990, 2004), particularly prominent in ESP (English for specific purposes), with the lexical bundles used in corpus research (Biber, Conrad, & Cortes, 2004; Biber, Johansson, Leech, Conrad, & Finegan, 1999). AWSuM auto-suggests high-frequency lexical bundles for each move within a section of a paper in a specific discipline. AWSuM (<http://langtest.jp/awsum/>) has been available free of charge since February 2016.

Figure 1 shows a screenshot of AWSuM. Compared with other reference tools, AWSuM includes the following advantages:

- It can display frequent lexical bundles within a section of a paper in a specific discipline according to rhetorical structures (i.e., moves).
- It can suggest the most frequent lexical bundles through an auto-complete user interface.
- It displays the words before and after the search terms, thus making it easier to check for collocations and formulaic sequences in a specific discipline.
- It includes a concordance feature, thus allowing it to display how words are actually used in the context of discipline-specific research articles.

Due to these features, AWSuM is significantly more useful than conventional reference tools such as dictionaries. We have also prepared a user's manual (http://langtest.jp/awsum/manual/AWSuM-Manual_E.pdf), which provides complete details about the features implemented in the tool.

Discipline
Applied Linguistics

Section
abstract

Move
02_presenting_research

Keyword
1 2 3 4 5 3R 4R 5R ☒ Auto-suggest

of this * was

- of this * was to investigate the
- of this * was to explore the
- of this * was to determine whether
- of this * was to examine the
- of this * was to examine whether
- of this * was to validate the
- of this * was to explore in
- of this * was to identify and
- of this * was to investigate a
- of this * was the point at

off 1L 2L

the aim

the objective

major purpose

the acquisition

Concordance

		"of this * was"	Search
The aim	of this study was	therefore to investigate the effectiveness of e	
The primary aim	of this study was	to assess the impact of a curricular intervent	
The goal	of this study was	to determine the overall effects of pronunciat	
The purpose	of this study was	to determine whether it is possible to disting	
The goal	of this study was	to determine whether learner beliefs regardin	

Google search

"of this * was" Search

Most frequent 4-grams in abstract 02_presenting_research

- as a foreign language
- this article reports on
- as a second language
- english as a second
- english as a foreign
- the extent to which
- this study investigated the
- this study examines the
- of english as a
- article reports on a
- this study investigates the
- this study examined the
- the purpose of this
- reports on a study
- of this study was
- learners of english as
- this study was to
- this article explores the
- of this article is
- this article is to
- this article examines the
- investigated the effects of
- this paper reports on

Figure 1. A Screenshot of the Web-Based Writing Support Tool "AWSuM."

From the beginning of its development, AWSuM has been based on the proof-of-concept of the bundle-move connection approach proposed by Cortes (2013). The bundle-move connection approach integrates two corpus-based approaches to the analysis and teaching of discourse structure: top-down and bottom-up (Biber, Connor, & Upton, 2007). The top-down approach focuses on macro-textual features such as moves. Thus, the analytical framework for the particular communicative functions (i.e., discourse units) is prepared first, and the analysis is then conducted based on the predetermined discourse unit types. On the contrary, the bottom-up approach focuses on linguistic features such as lexico-grammatical patterns. That is, the corpus investigation of vocabulary or grammar and grouping of its functions in texts comes first, and the discourse unit types are then identified based on linguistic criteria or groupings. In genre-based writing instruction, it is generally considered that it is the combination of a top-down approach to genre analysis and a bottom-up approach to corpus analysis that further helps raise learner awareness of rhetorical functions and linguistic features (Charles, 2007; Cortes, 2007; Cotos, 2009; Flowerdew, 2015).

Genre-based pedagogy is centered around rhetorical consciousness-raising (Hyland, 2004; Tardy, 2009), and technology can better address discipline-specificity and individualization of genre-based writing instruction (Cotos, 2014). Thus, in recent years, this line of corpus-based and genre-based writing instruction has been further

enhanced with the aid of technology (Anthony & Lashkia, 2003; Cotos, Huffman, & Link, 2015). AWSuM was designed to further this growth in technology-enhanced genre writing pedagogy. It is a data-driven and theory-based practical writing support tool for research articles (see Mizumoto, Hamatani, & Imao, in press, for details).

This article reports on a pilot study with L2 writers in Japan in order to evaluate its usability, effectiveness, and impact of AWSuM.

2 Method

AWSuM was pilot-tested with eight L2 writers in Japan: six undergraduate students (five females, one male) majoring in English who used the tool to write their Bachelor's theses in applied linguistics, and two graduate students (both females) writing their M.A. theses in applied linguistics. They participated in this small pilot study on a voluntary basis.

Their proficiency, based on their self-report of several proficiency test scores, was from B2 (independent user) to C1 (proficient user) in the Common European Framework of Reference (CEFR) for Languages (Council of Europe, 2001). Given their intermediate level of proficiency in English, the participants were considered appropriate as a target population sample (i.e., novice or non-native English researchers). None of them had written an RA prior to their thesis writing. Their theses followed the Abstract and Introduction, Methods, Results, Discussion, and Conclusion (IMRD/C) patterns, and consulting the developed tool as a reference resource would greatly help them with their linguistic difficulties. Thus, the participants' needs matched the purpose of the current pilot testing. No training was given to the participants except for explaining the concept of the tool.

The six undergraduates provided their feedback on an open-ended response sheet after using the developed tool for 2 months. The two graduate students were asked, for 5 months, to keep an online log of their search terms, open-ended comments, and feedback each time they used the tool while writing their manuscripts.

3 Results

The open-ended comments from all eight participants at the end of pilot testing came to a total of 284 entries, which were manually analyzed and classified into thematic categories. This analysis was conducted by the author of this article and a collaborative researcher, according to the following categories formulated from the literature on evaluating learners' responses to data-driven learning (DDL) tools (Mizumoto, Chujo, & Yokota, 2016; Yoon, 2011): (1) noticing of lexico-grammatical patterns (discovery), (2) confirmation of lexico-grammatical patterns (reference), (3) benefits unique to this tool, (4) greater confidence and autonomy in L2 RA writing, and (5) suggestions for improvement. The Kappa coefficient for the inter-rater reliability of the coding was 0.83, showing a very good level of agreement. Where disagreement occurred, a decision was made by consensus.

Excerpts of characteristic comments as well as overall frequencies for each category are provided in Table 1. The open-ended user comments suggest that the tool

Table 1. User Feedback on AWSuM (Categories, Frequency, and Excerpts)

---

(1) Noticing of lexico-grammatical patterns (Discovery) 67/284 entries (23.6%)	
-	I was able to learn the appropriate verb for a given noun. [e.g., excerpt # => excerpt # illustrates]
-	I noticed a number of verbs suitable for a certain context. [e.g., these studies have => these studies have shown]
-	I was able to learn the appropriate preposition for a given noun. [e.g., the Cronbach's alpha => the Cronbach's alpha for]
-	I was able to learn the specific nouns following a given adjective by setting the left-hand range. [e.g., has been a controversial => has been a controversial topic]
-	I was able to learn how technical terms are used. [e.g., Bonferroni => Bonferroni multiple comparison]
(2) Confirmation of lexico-grammatical patterns (Reference) 67/284 entries (23.6%)	
-	I was able to confirm that the phrase I came up with or wanted to use is frequently used. [e.g., procedure was followed]
-	I was able to confirm phrases for which I was unsure of their usage. [e.g., is defined as]
-	I was able to check if a phrase I had used in a different genre could be used in research articles. [e.g., the reason why]
-	I was able to check how an unfamiliar word like a synonym is used in research articles.
-	I was able to check how sentences could be constructed by viewing more suggestions.
(3) Benefits unique to this tool 58/284 entries (20.4%)	
-	Setting the section and the move helped me to find more appropriate phrases in the context. [e.g., university students who were majoring]
-	When I did not know how to begin a sentence, I looked at the "most frequent n-grams" in the move.
-	I came across a more appropriate phrase in the "most-frequent n-grams" when I set the section and move and tried to search for a different phrase.
-	This tool suggests frequently-used phrases for a given move, which made me realize how important those phrases are.
-	By fixing the left-hand range, I was able to find a more appropriate phrase in the context.
(4) Greater confidence and autonomy in L2 RA writing 39/284 entries (13.7%)	
-	I can write my paper with confidence because the tool suggests appropriate academic phrases.
-	I can write my paper with a more professional tone, compared with writing without any help.
-	As I use the tool, I may be able to learn many structural patterns and use them in my writing.
-	The more I used the tool, the more smoothly I was able to write. I want to keep using the tool.
-	I was able to feel assured that the phrases I had never used in my writing were actually used in other published papers.
(5) Suggestions for improvement 53/284 entries (18.7%)	
-	I felt I was not able to search for phrases well. It may take time to get used to using the tool adequately.
-	It would be better if more example sentences were suggested because I did not know how to begin a sentence.
-	Lexical bundles should be arranged and suggested according to their meanings.
-	It would be better if a dictionary or thesaurus were included in the tool.
-	While the concordance lines were useful, I was not sure which phrases would be most appropriate.

---

was useful for discovering and confirming lexico-grammatical patterns in research article moves in applied linguistics. Users also reported unique benefits of this tool, which most likely led to the development of greater confidence and autonomy in the users' L2 RA writing. These findings suggest that AWSuM indeed has the beneficial effects we set out to achieve in developing it. At the same time, users' feedback pointed to the difficulties they faced in making the most of using the tool (see "Suggestions for improvement" in Table 1). These difficulties will be addressed in future studies.

Taken together, although limited in terms of the number of participants, this pilot study provided valuable information about users' perceptions of the AWSuM's usability, effectiveness, and impact, as well as suggestions for its further improvement.

4 Conclusion

Corpus-based approaches to the analysis and teaching of discourse structure have increasingly become the established norm in language teaching and learning. One of the most advanced applications of such corpus-based approaches is their use in technology-enhanced genre L2 writing pedagogical tools (Anthony & Lashkia, 2003; Cotos, 2014). Against this background, AWSuM was designed to further the genre-based technology by developing a data-driven and theory-based practical writing support tool, which combines rhetorical moves and lexical bundles. The current pilot study showed that preliminary user feedback was mostly positive. Although larger-scale studies will be necessary to test the utility of AWSuM in detail, this pilot study confirmed that AWSuM could play a facilitating role in technology-enhanced genre writing teaching and learning.

Acknowledgements

This article is based on Mizumoto, Hamatani, and Imao (in press). This study was made possible by a Grant-in-aid for Scientific Research (No. 17H02369) from the Japan Society for the Promotion of Science. I would like to thank the anonymous reviewers of *Vocabulary Learning & Instruction* for their insightful comments and suggestions on the manuscript.

References

- Anthony, L., & Lashkia, G.V. (2003). Mover: A machine learning tool to assist in the reading and writing of technical papers. *IEEE Transactions on Professional Communication*, 46, 185–193. doi:10.1109/TPC.2003.816789
- Biber, D., Connor, U., & Upton, T.A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam, The Netherlands: John Benjamins.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405. doi:10.1093/applin/25.3.371
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, England: Longman.

- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6, 289–302. doi:10.1016/j.jeap.2007.09.009
- Cortes, V. (2007). Exploring genre and corpora in the English for academic writing class. *The ORTESOL Journal*, 25, 8–14.
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12, 33–43. doi:10.1016/j.jeap.2012.11.002
- Cotos, E. (2009). Designing an intelligent discourse evaluation tool: Theoretical, empirical, and technological considerations. In C.A. Chapelle, G.H. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 103–127). Ames, IA: Iowa State University. Retrieved from <https://www.apling.engl.iastate.edu/tsll/2008/12IADEElena.pdf>
- Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. New York, NY: Palgrave Macmillan.
- Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for EAP pedagogy. *Journal of English for Academic Purposes*, 19, 52–72. doi:10.1016/j.jeap.2015.05.004
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.
- Flowerdew, L. (2015). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes*, 20, 58–68. doi:10.1016/j.jeap.2015.06.001
- Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: University of Michigan Press.
- Mizumoto, A., Chujo, K., & Yokota, K. (2016). Development of a scale to measure learners' perceived preferences and benefits of data-driven learning. *ReCALL*, 28, 227–246. doi:10.1017/S0958344015000208
- Mizumoto, A., Hamatani, S., & Imao, Y. (in press). Bundle-move connection approach: Applying its concept to the development of an online writing support tool for research articles. *Language Learning*. Retrieved from <http://mizumot.com/files/LL-AWSuM.pdf>
- Swales, J.M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, England: Cambridge University Press.
- Swales, J.M. (2004). *Research genres: Explorations and applications*. Cambridge, England: Cambridge University Press.
- Tardy, C.M. (2009). *Building genre knowledge*. West Lafayette, IN: Parlor Press.
- Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10, 130–139. doi:10.1016/j.jeap.2011.03.003

Coverage-based Frequency Bands: A Proposal

Dale Brown

Kanazawa University

doi: <http://dx.doi.org/10.7820/vli.v06.2.Brown>

Abstract

Second language vocabulary research makes much use of word frequency lists and their division into bands. In recent years, bands of 1,000 items have become conventional. However, there does not seem to be any firm basis or rationale for this. Conventional banding may be questioned since the utility of words varies greatly depending on frequency, because there are enormous differences in frequency within higher bands, and because the reliability of the placement of words in bands becomes progressively poorer at lower frequency levels. This article suggests an alternative approach: basing bands on coverage levels. Because of the frequency distribution of words, this means the highest frequency bands would contain very few words, while lower frequency bands would contain a great many words. The article shows how such bands can be constructed and presents a re-analysis of the results of a vocabulary test designed with conventional bands in terms of coverage-based bands. This re-analysis produces a very different profile of learners' knowledge, and it is argued that the shape of this profile may be more useful in terms of guiding instruction in that it gives a clearer indication of which words should be targeted for a group of learners. It is further argued that the smaller number of words contained in coverage-based bands at higher frequency levels makes them a more feasible basis for instruction. The article thus concludes that coverage-based bands may be a fruitful avenue for researchers to explore.

1. Introduction

Much work on L2 vocabulary makes use of word frequency lists and in particular their division into bands. Bands of 1,000 items have become conventional through applications in vocabulary testing, such as Meara's yes/no tests (Meara & Milton, 2003; Meara & Miralpeix, 2006) and the vocabulary size test (Nation & Beglar, 2007), and the lexical profiling of texts (e.g., Nation, 2006; Webb & Rodgers, 2009) and of learners (Laufer & Nation, 1995; Milton, 2007).

There does not, however, appear to be any firm basis or an explicit rationale for the use of 1,000-item bands. Kremmel (2016) has recently criticized this convention since the utility of words varies greatly with frequency. This is clear in the coverage levels provided by successive bands. For example, Webb and Rodgers (2009) looked at coverage in a corpus of film scripts and found that while the first 1,000 word families covered 86.5% of the words in the corpus, the second 1,000 covered 4.2%, the third 1.7%, down to the tenth 1,000 which covered just 0.1% of the scripts. A second problem is that at higher frequency levels there is massive variation in the frequency of the items within a band. In Davies' COCA-based

(2008) lemmas list, for example, the most frequent lemma is 550 times more frequent than the 1,000th lemma, yet both, being in the first band, may be similarly regarded as high frequency items. By contrast, the 1,000th lemma is only 98 times as frequent as the 20,000th lemma. A third problem is that the reliability of the placement of words in particular bands is good at the highest frequency levels, but becomes progressively poorer for lower frequency bands. Comparing the first five 1,000-lemma bands from Davies with Kilgariff's (1995) similar BNC-based lemma bands, 83% of the lemmas overlap at the 1 K level, 61% at the 2 K level, 46% at the 3 K level, 34% at the 4 K level, and just 25% at the 5 K level. This problem is confirmed by Sorell's (2013) much more extensive data, while Miller and Biber (2015) raised concerns regarding the reliability of frequency rankings even when investigating a very narrow, focused corpus.

Kremmel's (2016) suggestion for reforming frequency banding was for bands of different widths at different frequency levels: 500-item bands at the highest frequency level, 1,000-item bands at mid-frequency levels, and 2,000-item bands at lower frequency levels. This article considers a more radical step in this direction and investigates whether the width of bands might be profitably based on coverage levels. Due to the frequency distribution of words, which is described by Zipf's (1935) law, this produces very small bands at the highest frequency levels and very wide bands at lower frequency levels. This is not an entirely new proposal: Edwards and Collins (2013), in developing a mathematical model of vocabulary acquisition, used bands based on the logarithmic function of a word frequency distribution, which has a similar outcome. There does not, however, seem to have been any previous discussion of how different approaches to banding may give us different perspectives on L2 vocabulary.

This article demonstrates how coverage-based bands may be constructed and considers how such bands may change our perceptions of vocabulary learning and teaching. This is achieved through the re-analysis of a data set from a yes/no test, which was constructed using conventional banding, on the basis of coverage-based bands.

2. Constructing Coverage-Based Frequency Bands

To construct coverage-based bands, Davies' COCA-based (2008) 20,000 lemmas list (which in fact includes a total of 20,200 lemmas) was used. The combined frequency of these lemmas is approximately 352 million, so, to create 10 bands, those at the top of the list were assigned to Band 1 until their combined frequency approximated 35.2 million. The next lemmas in the list were then assigned to Band 2 until approximately the same frequency was reached, and this process followed for each subsequent band. The result (Table 1) was that the highest bands are very narrow, while Band 10 is extremely wide. It was also not possible to divide the lemmas perfectly into 10 bands, Band 2 having a combined frequency somewhat larger than desired.

Next, a content-words-only approach was trialed. Gardner (2013) suggests that function words may be acquired differently from content words and notes they can distort frequency counts; Kremmel (2016) points out that most vocabulary tests do not include function words; while Milton (2009) suggests it can be useful at times for researchers to exclude them from frequency counts. Thus, all lemmas not listed

Table 1. Davies' Lemmas List Divided into 10 Bands Based on Coverage

Band	Lemmas included	Combined frequency of lemmas	No. of lemmas
1	1–2	34,584,731	2
2	3–6	38,225,595	4
3	7–15	35,052,835	9
4	16–35	35,165,890	20
5	36–83	35,017,770	48
6	84–210	34,979,487	127
7	211–514	35,034,325	304
8	515–1,243	35,010,971	729
9	1,244–3,315	35,001,760	2,072
10	3,316–20,200	33,922,071	16,885
Total		351,995,435	20,200

as nouns, verbs, adjectives, or adverbs were removed from the frequency list: a total of 446 (2.2%) of the lemmas. Ten bands were then formed with the 19,754 remaining lemmas, and this time the bands had very similar combined frequencies (Table 2).

These coverage-based bands directly deal with the first concern with conventional bands about differences in the utility of words of different frequency levels. They also address the second concern in that the items within each band are of more similar frequencies and there is a similar level of consistency across the bands. Finally, it is likely that coverage-based bands have greater reliability: with much wider bands at less frequent levels, about which there is less certainty of the frequency of items, it is more likely that most of the words would appear in the same band were bands constructed using another frequency list.

3. Using Coverage-Based Frequency Bands

One important use of frequency bands is in vocabulary testing, with a number of tests basing item selection on frequency bands. A test derived from

Table 2. Davies' Lemmas List, with Function Words Excluded, Divided into 10 Bands Based on Coverage

Band	Lemmas ranked	Combined frequency	No. of lemmas
1	1–3	19,423,367	3
2	4–27	19,305,146	24
3	28–88	19,583,571	61
4	89–207	19,561,491	119
5	208–412	19,590,257	205
6	413–767	19,557,796	355
7	768–1,375	19,564,642	608
8	1,376–2,525	19,559,456	1,150
9	2,526–5,275	19,555,439	2,750
10	5,276–19,754	19,652,327	14,479
Total		195,353,327	19,754

coverage-based bands could be designed in the same way. The vocabulary size test (Nation & Beglar, 2007), for example, uses a 1% sampling rate (i.e., one word on the test represents 100 words on the frequency list), while Meara's yes/no tests (Meara & Milton, 2003; Meara & Miralpeix, 2006) use a 2% sampling rate. Applying these rates to the coverage-based bands in Table 2, a test would require 201 items with a 1% sampling rate or 396 items with a 2% sampling rate (Table 3). These figures would, however, drop dramatically were Band 10 omitted. A test based on coverage-based bands is, therefore, quite feasible.

For the purposes of this article, it was unfortunately not possible to design and administer a test in this way. Instead, an existing data set was re-analyzed. These data were collected with a yes/no test, with items from Meara and Milton's (2003) X_Lex test and Meara and Miralpeix's (2006) Y_Lex test. The 240 items comprised 20 items each from 10 frequency bands, along with 40 pseudowords. Data were initially collected from 466 learners at four universities in Japan. Ninety-seven learners were excluded: 17 with an L1 other than Japanese, 10 since they had not fully completed the experimental tasks or were unwilling for their data to be used, and 70 with a false alarm rate over 25%. This left 369 learners in the study.

Figure 1 shows the mean percentage of words indicated as known in each of the 10 original frequency bands. This graph presents a familiar profile (e.g., Aizawa, 2006; Milton, 2007; Brown, 2012; Booth, 2013) with a gradual decline in the proportion of words known across the upper frequency bands. At lower frequency levels, there is more variability, which likely reflects the lower reliability of the placement of words in those bands.

The data were then re-analyzed by first finding the frequency ranking of each test item in Davies' lemmas list. The items were then assigned to the coverage-based bands in Table 2. Table 4 shows that the test contained no items at all from the first two bands, but had a reasonable sample of items from the other bands. Figure 2 thus shows the profile for the mean percentage of words indicated as known in the coverage-based bands. This shows high levels of knowledge (around ceiling level) up to Band 8, then a sharp drop at Band 9, and a further decline at Band 10.

Table 3. Sampling Coverage-Based Bands for a Vocabulary Test

Band	No. of lemmas	1% sample	Rounded (min. frequency = 1)	2% sample	Rounded (min. frequency = 1)
1	3	0.03	1	0.06	1
2	24	0.24	1	0.48	1
3	61	0.61	1	1.22	1
4	119	1.19	1	2.38	2
5	205	2.05	2	4.1	4
6	355	3.55	4	7.1	7
7	608	6.08	6	12.16	12
8	1,150	11.5	12	23	23
9	2,750	27.5	28	55	55
10	14,479	144.79	145	289.58	290
Total	19,754		201		396

Figure 1. Mean percentage of words known across the 10 original frequency bands.

Table 4. Sampling Rate of the Yes/No Test by Coverage-Based Frequency Bands

Band	No. of lemmas in band	No. of lemmas in yes/no test	Sampling rate (%)
1	3	0	0
2	24	0	0
3	61	3	4.9
4	119	4	3.4
5	205	10	4.9
6	355	9	2.5
7	608	19	3.1
8	1,150	17	1.5
9	2,750	35	1.3
10	14,479	91	0.6

A further analysis by deciles for both the 10 original bands and the 10 coverage-based bands is presented in Figures 3 and 4, respectively. Figure 3 shows that for each decile the approximately linear decline in the percentage of words known across the frequency bands is simply somewhat lower in each case. Figure 4 suggests that the pattern of ceiling levels of knowledge followed by a sharp drop seen in Figure 2 may begin at different points for different deciles.

These results suggest that coverage-based bands may be useful with regard to a second important application of word lists: guiding curriculum development and programs of instruction. The profiles in Figure 3 provide little guidance for instruction, it being unclear which of five or six bands should be targeted for any of the deciles. The profiles in Figure 4 seem somewhat more helpful. Although the results in this case are far from clear-cut, perhaps because many of the learners have already largely progressed through the majority of

Figure 2. Mean percentage of words known across 10 coverage-based frequency bands.

Figure 3. Mean percentage of words known across the 10 original frequency bands by deciles.

the bands and/or because the yes/no test was not designed specifically to investigate knowledge of these bands, a test derived from coverage-based bands may indicate more clearly the band at which learners' knowledge tails off. Coverage-based bands may, therefore, be of use diagnostically in identifying targets for instruction.

Figure 4. Mean percentage of words known across 10 coverage-based frequency bands by deciles.

Coverage-based bands may also be more practical with respect to instruction. Bands of 1,000-items are arguably too broad: providing instruction for 1,000 items is a considerable challenge within the typical horizon of a language course. Webb and Chang's (2012) study of Taiwanese learners over 5 years found typical growth of 200–300 word families a year, and they thus suggest a target of 400 families a year for a well-designed program. At this rate, it would take 2.5 years to progress through a single 1,000-family band. Coverage-based bands provide much smaller targets at the earlier stages of learning, which may well be more manageable and feasible (see also Dang and Webb [2016], who similarly argue that 1,000-item bands are impractical for pedagogical purposes and who divide their 800-item Essential Word List into 50-item bands to create readily attainable targets for teaching).

The corollary to small, manageable bands at high frequency levels is that at lower frequency levels the bands are much larger. These bands, of several thousand or many thousands of items, are clearly not manageable or feasible for explicit instruction. However, while there is debate about the exact point at which it should occur, it has been suggested that at some point in learners' development explicit teaching and learning should become secondary and implicit learning come to the fore (Nation, 2001; Schmitt, 2008). In addition, once learners have developed a good foundation of vocabulary, more specialized vocabulary appropriate to their learning purpose may be a more appropriate target for instruction (Nation, 2001). Thus, while the lowest bands may be infeasible for direct instruction, this may not be a major issue.

4. Conclusion

This article has proposed that dividing word frequency lists into bands based on coverage may have advantages over conventional 1,000-item bands. Conventional banding is questioned since words vary greatly in their utility depending on frequency, because such bands do not make the enormous differences in frequency within higher bands readily apparent, and because the reliability of the placement of words in bands becomes progressively poorer at lower frequency levels.

Basing bands on coverage produces upper bands that contain very few items, with the number of items in each band getting progressively bigger, until lower bands contain large numbers of items. Coverage-based bands deal directly with the issue of the variable utility of words, have smaller and more regular internal differences in frequency, and are likely to be more reliable in terms of the placement of words in bands.

A re-analysis of vocabulary test data in terms of coverage-based bands generated a very different profile of learners' knowledge, and it was argued that the shape of this profile may be more useful in guiding instruction. It was also argued that the smaller size of coverage-based bands at higher frequency levels makes them a more feasible basis for instruction.

Clearly the data presented here are very much preliminary. The bands constructed for this study, the method used, and the re-analysis of the vocabulary test data is intended to be illustrative. Word frequency lists and banding have multiple purposes and different approaches to banding may be appropriate for different purposes. Nevertheless, this article would like to argue that coverage-based bands are potentially informative and hopes that future research may make use of this proposal.

References

- Aizawa, K. (2006). Rethinking frequency markers for English-Japanese dictionaries. In S. Ishikawa, K. Minamide, M. Murata, & Y. Tono (Eds.), *English lexicography in Japan* (pp. 108–119). Tokyo: Taishukan Publishing Company.
- Booth, P. (2013). Vocabulary knowledge in relation to memory and analysis: An approximate replication of Milton's (2007) study on lexical profiles and learning style. *Language Teaching*, 46(3), 335–354. doi:10.1017/S0261444813000049
- Brown, D. (2012). The frequency model of vocabulary learning and Japanese learners. *Vocabulary Learning and Instruction*, 1(1), 20–28. doi: 10.7820/vli.v01.1.brown
- Dang, T.N.Y., & Webb, S. (2016). Making an essential word list for beginners. In I.S.P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153–167). Amsterdam: John Benjamins.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present*. Retrieved from <http://www.american corpus.org>

- Edwards, R., & Collins, L. (2013). Modelling L2 vocabulary learning. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 157–183). Amsterdam: John Benjamins.
- Gardner, D. (2013). *Exploring vocabulary*. London: Routledge.
- Kilgarriff, A. (1995). *BNC database and word frequency lists*. Retrieved from <http://www.kilgariff.co.uk>
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987. doi: 10.1002/tesq.329
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322.
- Meara, P., & Milton, J. (2003). *X_Lex, The Swansea Levels Test*. Newbury: Express.
- Meara, P., & Miralpeix, I. (2006). *Y_Lex: The Swansea Advanced Vocabulary Test. v2.05*. Swansea: Lognostics.
- Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20(1), 30–53. doi:10.1075/ijcl.20.1.02mil
- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modeling and assessing vocabulary knowledge* (pp. 47–58). Cambridge: Cambridge University Press.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- 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? *The Canadian Modern Language Review/La revue canadienne des langues vivantes*, 63(1), 59–81. doi: 10.3138/cmlr.63.1.59
- Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. doi:10.1177/1362168808089921
- Sorell, C.J. (2013). *A study of issues and techniques for creating core vocabulary lists for English as an international language*. Doctoral dissertation, Victoria University of Wellington.
- Webb, S., & Rodgers, M.P.H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407–427. doi: 10.1093/applin/amp010
- Webb, S.A., & Chang, A.C.-S. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126. doi: 10.1177/0033688212439367
- Zipf, G.K. (1935). *The psycho-biology of language*. Boston, MA: Houghton-Mifflin.