English Language Teaching; Vol. 12, No. 10; 2019
ISSN 1916-4742 E-ISSN 1916-4750
Published by Canadian Center of Science and Education

Developing a Technical Words List for Research Articles in Computer Science Discipline

Haiyuan Chen[1] & Gang Lei[2]

[1] Foreign Languages College, Jiangxi Normal University, Nanchang, China

[2] School of Software, Jiangxi Normal University, Nanchang, China

Correspondence: Haiyuan Chen, Foreign Languages College, Jiangxi Normal University, Nanchang, China.

Received: August 6, 2019 Accepted: September 26, 2019 Online Published: September 28, 2019

doi: 10.5539/elt.v12n10p131 URL: https://doi.org/10.5539/elt.v12n10p131

Abstract

This study aims to identify technical words in research articles across sub-disciplines in computer science field. A corpus of research articles in computer science was constructed, comprised of 10,450,035 running words from 1045 research articles across ten sub-disciplines of computer science. The technical word List for research articles in computer science was created by the methods of computer-based analysis and the rating scales. 769 technical word types were identified and listed in Appendix A. The computer science field-oriented technical words list is expected to meet the needs of learners in this field.

Keywords: technical words, research articles, computer science discipline, computer-based analysis, rating scales

1. Introduction

According to Nation (2001), English words can be classified into four categories: high-frequency words; academic words; technical words and low-frequency words. The nature and coverage of high frequency and academic words have been attached great importance. The word lists had been developed such as General Words List (West, 1953), the new General Words List (Brezina & Gablasova, 2015), a new Academic Words List (Coxhead, 2000) and the new Academic Vocabulary List (Gardner & Davies, 2014). The discipline-based academic wordlists in certain fields had also been developed, such as medicine (Wang, Liang, & Ge, 2008), agriculture (Martínez, Beck, & Panza, 2009), applied linguistics (Vongpumivitch, Huang, & Chang, 2009), business (Chen, Hu, & Ho, 2009), engineer (Ward, 2009), chemistry (Valipouri & Nassaji, 2013), and nursing (Yang, 2015). However, these discipline-based wordlists cover a large number of academic words, but the technical words strongly related to the specific disciplines do not account for the most part. For example, a nursing academic word list (NAWL) created by Yang (2015) includes 676 word families, out of which, 378 word families overlap with the AWL. The ten most frequent words in the NAWL are *participate, significant, data, research, clinic, analyze, assess, score, respond, symptom*. Only the two words *clinic* and *symptom* are related to the nursing field.

Technical words are useful to people in a specialized field. It is necessary and important to provide students with a more restricted and discipline-based wordlist (Hyland & Tse, 2007). Recently, researchers compared different ways to identify technical words (Kwary, 2011; Chung & Nation, 2003, 2004) and developed the technical words list in the fields of trades and finance (Coxhead & Demecheleer, 2018; Patanasorn, 2018).

The researchers and graduate students in China have to read research articles published in English. Due to lack of mastery of the technical words in specialized field, they still feel distressed in reading the research articles in English. The study aims to identify the technical words in research articles in the computer science discipline. To my knowledge covered, no list has exclusively targeted research articles in the computer science discipline.

2. Literature Review

2.1 Technical Words

Technical words occurred frequently in a specialized text or subject area (Nation, 2001). For example, it is difficult for people who are not in the field of computer science to know the words such as *boolean, petri, and*

snippet. Some technical words now become the everyday words in English, such as *folder, database and disk* in computer science. Some every day words may have the technical meaning, such as *query, recall and server*.

Technical words can account for a certain proportion of the words in the specialized texts. Chung and Nation (2003 and 2004) stated that technical words constitute more than 30% of tokens in medical texts. Patanasorn (2018) developed a frequent technical word lists for finance, containing 979 technical words (569 headwords and their word family members, 413words appear in the GSL, 291words appear in the AWL wordlist and 275 words are unique to the finance field. Coxhead and Demecheleer (2018) investigated the technical vocabulary of plumbing, which covers over30% of the written corpus and over 11% of the spoken corpus.

2.2 Identifying Technical Words

There are several methods used to identify technical words: rating scales (Chung & Nation, 2003); a technical dictionary (Oh et al., 2000); context clues (Flowerdew, 1992); computer-based method (Chung & Nation, 2004; Anthony, 2010) and hybrid method (Kwary, 2011; Patanasorn, 2018; Coxhead & Demecheleer, 2018). Having compared the first four different methods of identifying technical words, Chung and Nation (2004) concluded the rating scale method is the most accurate but time-consuming and the computer-based method works best in terms of practicality.

Chung and Nation (2003) designed a four-step rating scale to identify technical words. In Step1, experts identify words with no particular relationship with a specialized area. For example the function words like the, is, by would be classified as the least technical words. In Step 2, experts tell words whose meaning is minimally related to a specialized area. In Step 3, experts identify words with meaning closely related to a specialized area. In Step 4, experts identify words whose meaning is specific to a specialized area. The context is of great value for experts to identify technical words or non-technical words in the four steps.

Computer-based method can be done in two ways, by using vocabulary classification program like Range program and by using keyword analysis program like AntConc. The Range program (Healty et al., 2002) provides a range figure (how many texts the word occurs in) and a headword frequency figure (the total number of times the actual headword type appears in all the texts). Three word lists (the 1000, 2000 high frequent words lists and a 570 academic words list) in the Range program were used as stop lists. Words in the stop lists were excluded from the new list created by computer automatically. Running the Range program generated Type and Family output, which provides information on the overall coverage of the four lists (the 1000, 2000 high frequent words lists and the 570 academic words list and the Not Found in Any List). The Not Found in Any List covers technical words and low frequency words. The keyword analysis program AntConc (Anthony, 2010) generated keywords. A keyword means a word which occurs with unusual frequency in a given text (Scott, 1997). The technical texts should be large enough to ensure that there are plenty opportunities for the technical words to occur. The non-technical reference corpus is not allowed to include any texts from the target corpus and it should not be smaller than the target corpus. Kwary (2011) stated that the Range program causes a problem: some words in high frequency words lists may also belong to technical words, but they are excluded due to the stop lists. There are also weak points in AntConc: the absence of the word class and the production of only single-word units.

Realizing the drawbacks of a single method, researchers began to resort to hybrid method to identify the technical words. Kwary (2011) mixed the keyword analysis method and the systematic classification method to identify the technical words in a financial text. Patanasorn (2018) combined keyword analysis method with a modified rating scale to develop a frequent technical words list for finance. Coxhead and Demecheleer (2018) used the Range program and rating scale method to investigate the technical vocabulary of plumbing.

These studies shed light on the necessity and practicality of developing technical words list for a specialized area. Considering the students' need of technical words for reading research articles in computer science discipline, this paper focused on developing a list of frequent technical words in corpus of research articles in computer science(hereafter CRACS)via a combined method. The Range program and the rating scale will be combined to identify the technical words in CRACS. The following two questions will be answered in this paper.

1) How many technical words are there in CRACS?
2) How are the technical words in CRACS distributed in terms of range and frequency?

3. Methodology

The CRACS was comprised of 10,450,035 running words from 1045 computer science research articles across ten sub-disciplines of computer science. This corpus-based study aims to identify technical words in CRACS.

3.1 The Corpus

The construction of CRACS followed the criteria proposed by Sinclair (2005) covering representatives, specificity of corpus, use of whole documents and availability in electronic form. With the subject specialist's recommendation in our university, we extract research articles published in journals across the ten sub-disciplines, which are shown in Table 1 below. CRACS contains 1045 articles from ten journals published during 2010-2014. The large capacity of the CRACS supports Chung and Nation's (2003) suggestion that larger representative corpus of a technical field would be needed for listing a definitive technical vocabulary for that field. The corpus represents a genre, the experimental research articles, and a field, the computer science. The articles are stored in the form of texts with their reference, appendices, captions, footnotes and acknowledgements removed, keeping the introduction-method-results-discussion format.

3.2 The Computational Analysis

Chung and Nation (2004) compared four different approaches to identifying technical words in an anatomy text and concluded that computer-based approach works the best in terms of practicality. The computer software Range (Heatly et al., 2002) was employed for the analysis. The output provided information on the overall coverage of the four lists (the first and second thousand words of the GSL, the AWL and Not found in any list). The Not Found in Any List covers technical words and low frequency words. The computational result was shown in Table 2.

The three criteria including range, frequency and word type were used to create a computer science technical word list. Coxhead (2000) argued that the range threshold can ensure that words count are found across the breadth of the texts rather than related to particular longer texts or topic related words. Words had to appear in at least 5 out of the10 sub-disciplines. The frequency threshold in past studies ranged between 10 and 40 times per million tokens. As far as the 10 million words in the present corpus were concerned, appearing 100 times at least was the selection thresholds. Word type is single word like *alter or alters*. Bauer and Nation (1993) defined that a word family includes the basic word plus inflected forms and transparent derivations, for example *ALTER: alter, alterable, alteration, alterations, altered, altering, alternate, alternating, alters, unalterable and unaltered*. Chung and Nation(2003) argued that word types, rather than word families, were used as the unit of counting the technical vocabulary in specialized texts on the ground that one or two members of a family were technical words, not all of them were technical words. Therefore this study followed word type criterion to count words. With the help of the Range program, 1262 word types were extracted by adopting the criteria of range, frequency and type.

Table 1. Sub-disciplines of computer science

No. sub-discipline	No. sub-discipline
1. The computer system & high performance computing	6. The computer network
2. Database / /content retrieval data mining	7. Computer science theory
3. The artificial intelligence & pattern recognition	8. Network & information security
4. Human-computer interaction &pervasive computing	9. Computer graphics &multimedia
5. Software engineering/programming language	10. Front/cross/composite

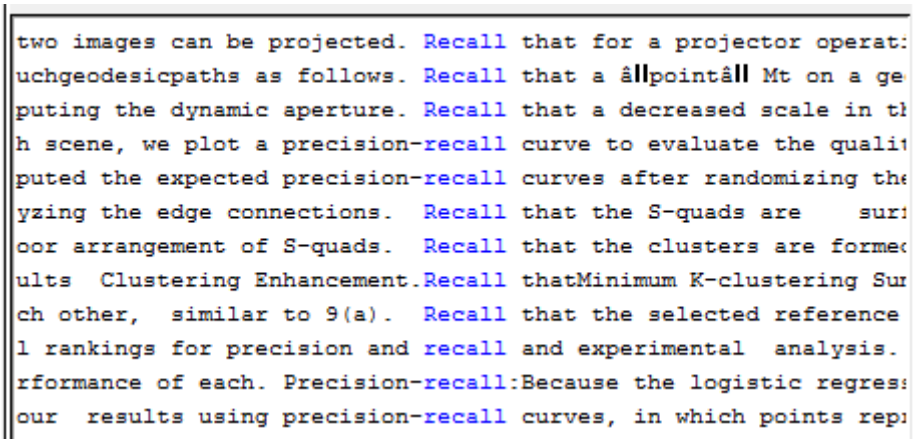
Table 2. Word types in CRACS

Word list / Types/%	First 1000	second 1000	Academic word	Not in the list	Total
	3385/1.76%	2888/1.50%	2807/1.46%	183613/95.29%	192693/100%

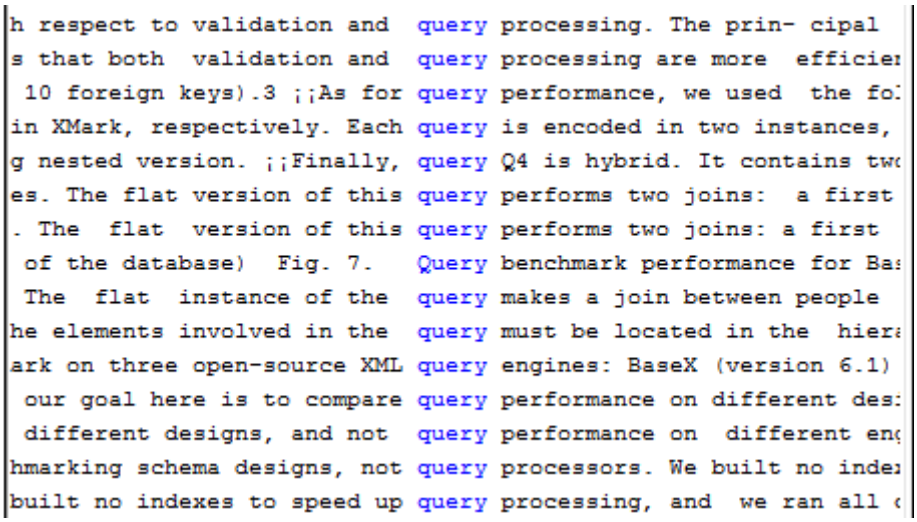
3.3 Manual Refinement

This list of 1262 word types required further scrutiny in order to select and present the technical words in a more systematic and pedagogically useful way. The 1262 word types underwent the four-step rating scales. Each word was assessed independently by the two researchers to determine whether a specific word should be included or not. In step 1, those words not related to computer science but used in general language were excluded. In step 2 words with minimally related to a computer science were excluded. In Step 3, words with meaning closely related to computer science were included. In Step 4, words specific to computer science were included. With regard to their advantages of the more current, accurate and comprehensive lists, the new general service list

(new GSL) (Brezina & Gablasova, 2015) and the academic vocabulary list (AVL) (Gardner & Davies, 2014) were applied to identifying technical words from the 1262 word types. Therefore, the aim of manual vetting is to identify those GSL, new GSL, AWL, and AVL words with high frequency and specific meanings in computer science discipline. Word like *recall* in the new GSL and word like *query* in the AVL should not be excluded from the technical words list when their specific meanings, uses and collocations are taken into considerations. Chung and Nation (2004) proved that if common collocations of terms are included to identify the technical words, the rate of correct identification of technical terms will rise. Picture 1 and Picture 2 show that the word *recall* is highly collocated with *precision* and *query* is highly collocated with *processing* and *performance.* Words *recall* and *query* in the CRACS have their specific meanings.


two images can be projected. Recall that for a projector operat:
uchgeodesicpaths as follows. Recall that a âllpointâll Mt on a ge
puting the dynamic aperture. Recall that a decreased scale in tl
h scene, we plot a precision-recall curve to evaluate the qualit
puted the expected precision-recall curves after randomizing the
yzing the edge connections. Recall that the S-quads are sur:
oor arrangement of S-quads. Recall that the clusters are formec
ults Clustering Enhancement.Recall thatMinimum K-clustering Sur
ch other, similar to 9(a). Recall that the selected reference
l rankings for precision and recall and experimental analysis.
rformance of each. Precision-recall:Because the logistic regres:
our results using precision-recall curves, in which points rep:

Picture 1. Collocations of *recall*


h respect to validation and query processing. The prin- cipal
s that both validation and query processing are more efficie:
10 foreign keys).3 ;;As for query performance, we used the fo:
in XMark, respectively. Each query is encoded in two instances,
g nested version. ;;Finally, query Q4 is hybrid. It contains tw
es. The flat version of this query performs two joins: a first
. The flat version of this query performs two joins: a first
of the database) Fig. 7. Query benchmark performance for Ba:
The flat instance of the query makes a join between people
he elements involved in the query must be located in the hier:
ark on three open-source XML query engines: BaseX (version 6.1)
our goal here is to compare query performance on different des:
different designs, and not query performance on different en
hmarking schema designs, not query processors. We built no inde:
built no indexes to speed up query processing, and we ran all

Picture 2. Collocations of *query*

4. Results and Discussion

4.1 Technical Word List for Research Articles in Computer Science

After the stages of computational analyses and manual refinement as elaborated above, the Technical Word List for Research Articles in Computer Science (hereafter TWLRACS) with 769 word types was constructed. The percentage of words in the four lists was presented in Table 2. And it shows that words in the Not Found in Any List make up 95.29% of the total types, a most striking coverage by the technical words and low frequency words. 1262 word types were extracted from the Not Found in Any List which met the range and frequency standards (range ≥ 5; frequency ≥ 100). There are 769 word types left after the manual refinement and there are 333 AVL words, 50 new GSL words and 110 other words excluded from the 1262 word types. The new GSL and particularly the AVL are helpful in filtering non-technical words, which overcoming the drawbacks of West's

GSL and Coxhead's AWL as baseword lists in the range software. The top ten technical words are presented in Table4. Words like *algorithm, node, query* covered all the ten sub-disciplines areas and occurred more than 10,000 times. The top ten AVL words excluded from the 1262 word types were presented in Table 5. AVL words like *graph, linear, matrix, mapping, spatial, static, metric, temporal, partition and calculus* covered all the 10 sub-disciplines areas and occurred more than 1000 times. The top ten new GSL words excluded from the 1262 entries were shown in the Table 6. The new GSL words such as *traffic, client, reference, candidate, score, capture, mobile, robust, cell, clients* also had high coverage and frequency. Though the 333 AVL words and 50 new GSL words do not belong to technical words, their high coverage and frequency alerted that there is no reason to ignore their significance. They are listed in Appendix B and Appendix C respectively.

Table 3. Types of filtered vocabulary in 1262 entries

technical words (Pre-filtered)	*words in the new GSL*	*words in the AVL*	*other words (eg.persons' name)*	*Technical words(Pra-filtered)*
1262	50	333	110	769

Table 4. Top ten technical words in TWLRACS

Words	*Range*	*Frequency*
ALGORITHM	10	20084
NODE	10	13452
QUERY	10	11607
THEOREM	10	7893
LEMMA	9	6927
DATABASE	10	5043
SEMANTICS	10	4043
PACKET	7	3891
KERNEL	10	3508
TUPLES	8	3061

Table 5. Top ten words occurred in the AVL

Words	*Range*	*Frequency*
GRAPH	10	6841
LINEAR	10	4891
MATRIX	10	4330
MAPPING	10	2861
SPATIAL	10	2675
STATIC	10	2175
METRIC	10	1902
TEMPORAL	10	1842
PARTITION	10	1730
CALCULUS	10	1596

Table 6. Top ten words occurred in the new GSL

Words	*Range*	*Frequency*
TRAFFIC	7	3327
CLIENT	9	2492
REFERENCE	10	2020
CANDIDATE	10	1922
SCORE	10	1799
CAPTURE	10	1684
MOBILE	10	1620
ROBUST	10	1574
CELL	10	1570
CLIENTS	7	1498

4.2 Distribution in Terms of Range and Frequency

This part analyzed the distribution of the 769 technical word types in the TWLRACS in terms of range and frequency. Table 7 shows that how the 769 technical word types in computer science had been distributed in terms of range. Of the 769 technical words in the list, 202(26.27%) of word types covered all the 10 subject areas, 147(19.11%) covered 9 subject areas, 116(15.08%) covered 8 subject areas, 154(20.03%) covered 7subject areas, 98(12.74%) covered 6 subject areas and 52(6.76%) covered 5 subject areas. More than half of the words (465) occurred in 8 or more of the 10 sub-disciplines areas. Table 8 presents us the frequency distribution of the 769 technical words in the TWLRACS. Of the 769 technical words in the list, 29 word types (3.77%) occurred more than 2000 times in the CRACS. 55 word types (7.15%) occurred more than 1000 times but fewer than 2000 times. There are 305 word types with frequency ranging from100 to 199 times, which accounts to 39.66.69% of the total technical word types. More than half of the words (438) in the TWLRACS appeared with frequency ranging from 100 to 299. The clear picture of the frequency distribution can be described as the following: the higher the frequency is, the less word types were distributed. Table 9 lists the top 20 technical words which are with frequency more than 2000 times and coverage of the 10 sub-disciplines. Richards (2001:7) claims that“ words with the highest- frequency and the widest range are considered to be the most useful ones for the purpose of language teaching” .Therefore, the 769 technical word types in the TWLRACS were listed by their range and frequency order instead of by alphabetic order.

Table 7. Subject area of coverage of 769 in the TWLRACS

Subject area covered	*Number of word types*	*percentage*	*Accumulative percentage*
10	202	26.27%	26.27%
9	147	19.11%	45.38%
8	116	15.08 %	60.46%
7	154	20.03%	80.49%
6	98	12..74%	93.24%
5	52	6.76%	100%
Total	769	100%	

Table 8. TWLRACS word distribution in the CRACS

Times of occurrence	*Number of words*	*percentage*	*Accumulative percentage*
>2000	29	3.77%	3.77%
1000-1999	55	7.15%	10.92%
500-999	108	14.04%	24.96%
300-499	139	18.08%	43.04%
200-299	133	17.30%	60.34%
100-199	305	39.66%	100%
Total	769	100%	

Table 9. Top 10 technical words (with10 ranges and more than 2000 times of occurrence)

Words	*Range*	*Frequency*
ALGORITHM	10	20084
NODE	10	13452
QUERY	10	11607
THEOREM	10	7893
DATABASE	10	5043
SEMANTICS	10	4043
KERNEL	10	3508
PROPOSITION	10	3037
SERVER	10	2941
BINARY	10	2773

5. Conclusion

The present study focused on extracting technical words in research articles of computer science discipline and it provided a list of 769 technical word types with high frequency and wide range. The list is based on the computational analyses of ten sub-disciplines research journals in the computer science field and manual refinement of the 1262 word types. Such computer science field-oriented list will more closely meet the needs of learners in this field. As Durant (2014) proved, disciplines appear to be good units of analysis for vocabulary listing. The present study also produced a list of 333 AVL word types that are frequently used in research articles of computer science discipline. Gardner and Davies (2014) believed that the AVL is the most current, accurate and comprehensive list of core academic vocabulary in existence today. The AVL functioned as a useful filter in helping us to exclude non-technical words. Furthermore, our result lend support to the argument that some GSL and AWL words are used with academic meaning or technical meaning in specialized contexts (Martinez et al., 2009).

The future research will attempt to develop the collocation of the 769 technical word types in CRACS. Chung and Nation (2003) pointed that the common collocations and grammatical patterns of the technical words were beneficial to students.

Acknowledgements

This study is financially supported by the Jiangxi Provincial Education Reform Project (Grant No. JXJG-17-2-36).

References

Anthony, L. (2010). AntConc 3.2.1 http://www.antlab.sci.waseda.ac.jp/software.html

Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics, 36*(1), 1-22. https://doi.org/10.1093/applin/amt018

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279.

https://doi.org/10.1093/ijl/6.4.253

Chen, P., Hu, K., & Ho, J. (2009). A study of academic vocabulary used in the abstracts of business and management. *Taiwan International ESP Journal, 1*, 51-76.

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-115.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System, 32*, 251-263. https://doi.org/10.1016/j.system.2003.11.008

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-23. https://doi.org/10.2307/3587951

Coxhead, A., Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes, 51*, 84-97. https://doi.org/10.1016/j.esp.2018.03.006

Durrant, P. (2014). Discipline and Level Specificity in University Students' WrittenVocabulary. *Applied Linguistics, 35*(3), 328-356. https://doi.org/10.1093/applin/amt016

Flowerdew, J. (1992). Definition in science lectures. *Applied Linguistics, 13*(2), 202-221. https://doi.org/10.1093/applin/13.2.202

Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics, 35*(3), 305-327. https://doi.org/10.1093/applin/amt015

Heatley, A., Nation, P., & Coxhead, A. (2002). Range Program. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*(2), 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Kwary, D. (2011). A hybrid method for determining technical vocabulary. *System, 39*, 175-185. https://doi.org/10.1016/j.system.2011.04.003

Oh, J., Lee, J., Lee, K., & Choi, K. (2000). Japanese term extraction using dictionary hierarchy and a machine translation system. *Terminology, 6*, 287-311. https://doi.org/10.1075/term.6.2.09oh

Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes, 51*, 45-54. https://doi.org/10.1016/j.esp.2018.03.002

Martínez, I. A., Beck, S. C. & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 28, 183-198. https://doi.org/10.1016/j.esp.2009.04.003

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Scott, M. (1997). PC analysis of keywords and key keywords. *System, 25*(2), 233-245. https://doi.org/10.1016/S0346-251X(97)00011-0

Sinclair, J. (2005). Corpus and text – Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxford: Oxbow Books.

Swales, J. M. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.

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

Vongpumivitch, V., Huang, J., & Chang Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*, 33-41. https://doi.org/10.1016/j.esp.2008.08.003

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical word list. *English for Specific Purposes, 27*, 442-458. https://doi.org/10.1016/j.esp.2008.05.003

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*, 170-182. https://doi.org/10.1016/j.esp.2009.04.001

West, M. (1953). *A General Service List of English Words*. Longman, London.

Yang, M. (2015). A nursing academic word list. *English for Specific Purposes, 37*, 27-38. https://doi.org/

10.1016/j.esp.2014.05.003

Appendix

Appendix A (769 technical words)

1.TYPE2.ALGORITHM3.NODE4.QUERY5.THEOREM6.DATABASE7.SEMANTICS8.KERNEL9.PROPOSI-TION10.SERVER11.BINARY12.TUPLE13.INTERFACE14.THRESHOLD15.VERTICES 16.BANDWIDTH17.RECALL18.CLUSTER19.SUBSET20.VERTEX21.BUFFER22.STORAGE23.TOPOLOGY24.SOFTWARE25.SEGMENTATION26.HARDWARE27.POLYNOMIAL28.LOOP29.RU NTIME30.STACK31.FEEDBACK32.WIRELESS33.SEMANTIC34.QUEUE35.INTEGER36.BOOLEAN37.SEGMENT38.SENSOR39.SPARSE40.GRID41.ATOMIC42.GAUSSIAN43.NOTATION44.W ORKLOAD 45.SEGMENTS 46. ARCHITECTURE 47. SESSION 48. INTERNET 49.DISK 50.RECURSIVE 51.VERIFY 52.CORRELATION 53.DATASETS 54. DESCRIPTOR 55.MATRICES 56.DIGITAL 57.FILTER 58. DISJOINT 59.PAIRWISE60.COMPUTES61.MODULE62.CONCRETE63.SETUP64.BENCHMARK65.IMPLEMENTATIONS

66.HTTP67.ALGEBRA68.COUNTER69.REWRITE70.DUAL71.EXECUTE72.DEFAULT73.ROTATION74.EUCLIDEAN75.PORT76.PROTOTYPE77.TAG78.NONLINEAR79.VER-TICAL80.LOOKU P81.TRIANGLE82.MAGNITUDE83.TRIVIAL84.HORIZONTAL85.LAYOUT86.SCAN87.INCOMING88.OPTIMIZE89.REUSE90.SUBGRAPH91.DENSE92.ORTHOGONAL93.COMPACT94.TECH NOLOGIES95.ASYNCHRONOUS96.C-ARDINALITY 97. ENTROPY 98.KEYWORD 99.PERMUTATION100.INDICES101.INTEL102.CORRELATED103.HISTOGRAMS104.QUADRATIC105.PENALTY106.SOLVER107.TILE108.GRANULARITY109.SCALABLE110.COMPRESSE-D111.PREPROCESSING 112.OUTGOING 113.BIN114.GRAY115.GOOGLE116.MATHEMATICAL117.HOMOGENEOUS118.ARITHMETIC119.ARCHITECTURES120.SKETCH 121.TRIPLES 122.MASK 123.NAIVE 124.MEDIAN 125.SNAPSHOT 126. SLICING 127. INSTANTIATION128.DASHED129.OPTIMUM130.NONZERO131.BATCH132.QUANTIFY133.OUTPERFORM134.PRIORI135.CORRELATIONS136.ROUTINE137.RESIDUAL138.RETRIEVE139. DEDICATED140.ENDP-OINT141.DUPLICATE142.LOGARITHMIC143.SKIP144.COMPETITIVE145.SEMANTICALLY146.ANONYMOUS147.EXIT148.DESKTOP149.PRUNE150.TRACTABLE1 51.RECTANGLE152.HTML153.TRAVERSE154.ALPHA155.SUBSECTION156.PARSE157.MATION158.SUBTLE159.SUITE160.INTERSECT161.ASYMPTOTICALLY162.COPE163.LABORATOR-Y164.POP165.FEED166.SUBOPTIMAL167.DRAWBACK168.APP169.DISCARD170.SUPER171.DIAMETER172.COMPROMISE173.DECOMPOSE174.ANOMALIES175.RESET176.NONUNIFOR1 77.CARTESIAN178.WARD179.RIGHTMOST180.DOWNLOAD181.REVISIT182.ACCOMPLISH183.DOTTED184.MULTIPLICATIVE185.IMPRACTICAL186.ALARM187.DISCREPANCY188.DIS CONNECTE189.DIAGNOSIS190.MISMATCH191.DEPLOY192.SIBLINGS193.DEPICT194.DRASTICA-LLY195.INDISTINGUISHABLE196.MIRROR197.NONLINEARITY198.INTERLEAVED199 .LAPTOP200.ENSEMBLE201.TEDIOUS202.DENOMINATOR203.UBIQUITOUS204.LEMMA205.VIDEO206.THROUGHPUT207.CACHE208.SCHEMA209.C-LUSTERING210.PIXEL211.DATASE T212.HASH213.MESH214.DESCRIPTORS215.SERVERS216.DETERMINISTIC217.USABILITY218.REWRITING219.BASELINE220.CANONICAL221.SYNTAX222.CONVEX223.LINUX224.SYN THETIC225.CONFIDENTIALITY226.PEDESTRIANS227.TOKEN228.FILTERING229.COMPILER230.PATCHES.231.ORACLE225.HISTOGRAM226.BYTES232.PIPELINE233.KERNELS234.CHU NK235.MARKOV236.MODAL237.MAC238.FILTERS239.SENSORS240.SPANNING241.REMOTE242.RECURSION243.SLICE244.POSTERIOR245.STOCHASTIC246.COLLISION247.ACYCLIC2 48.OUTLIERS249.THEOREMS250.NOMINAL251.RELAY252.NOTATIONS253.NONTRIVIAL254.OPTIMIZING255.EXECUTIONS256.CONTRACTION257.LA-TTICE258.BINS259.TOPOLOGIC AL260.BAYESIAN261.CIRCUIT262.SIBLING263.INVARIANTS264.ALGEBRAIC265.TRIPL-E266.SCALAR267.PLACEMENT268.PENDING269.MONOTONIC270.TRIVIALLY271.PLATFORMS 272.REWRITTEN273.SQUARED274.COMPOSITIONS275.CORPUS276.ARCHITECTURAL277.ASYMPTOTIC278.UNLABELED279.INSERTIONS280.SUBSEQUENCE281.PARAMETERIZATIO N282.TILES283.PREDECESSOR284.SUBSTRINGS285.ADMISSIBLE286.SATURATION287.MODULAR288.SKETCHES289.DELTA290.COLLISIONS291.UNDIRECTED292.METADATA293.DRI FT294.LEVERAGE295.INTERIOR296.MONOTONICALLY297.RECTANGULAR298.CLUSTERED299.DUMMY300.DAG301.OUTLIER302.MULTIVARIATE303.TIVELY304.PSEUDOCODE305.S LICES306.SUBGRAPHS308.BIDIRECTIONAL309.FACTORIZATION310.TAGGING311.INSTALLED312.SERIAL313.TRIANGULAR314.HINTS315.AMORTIZED316.TRAVERSING317.SCANNE D318.PIPELINES319.PREDICTIVE320.SWAP321.MACRO322.SATURATED323.QUAD324.INSTANTIATE325.EXPONENT326.CALIBRATED327.PARALLELIZATION328.STACKED329.OUTPE RFORMED330.DOWNLOADED331.MULTILEVEL332. OVERLOAD333. ORDERINGS334. ABSTRACTED 335. SUBCLASSES336 BLANK337. OPAQUE338. SOLVERS339. PRONE340. DEPARTURE341. GPS342. MITIGATE343. SUPERSET344. LEGACY345. LATED346. PHYSICS347.SKEW348.RECALLING349.ANCESTORS350.PROTOTYPES351.ALLEVIATE352.TUPLES353.PIXELS354.LATENCY355.AUTOMATA356.PATCH357.BLUR358. ALPHABET359. SLOTS360. SUBSPACE361. UNARY362. SPEEDUP363.REACHABLE364.CONFIGURATIONS365.DIMENSIONALITY366.RADIUS367.SYNTACTIC368.COVARIANCE369.COEFFICIENTS370.BUFFERS371.LEMMAS372.INDUCTIVE3 73.AMBIENT374.SPECTRAL375.COEFFICIENT376.BYTE377.OFFLINE378.ISOMORPHIC379.OVERLAY380.PLANAR381.AUDIO382BITMAP383.PROPOSITIONS384.QUEUES385.SENSING3 86.PARAMETRIC387.HASHING388.MONOTONE389.ANIMATION390.INJECTION391.ANCESTOR392.TRIANGLES393.MODALITIES394.TOPOLOGIES395.REORDERING396.SNIPPET397.E XECUTABLE398.POS399.MODULO400.SUBSTRING401.HULL402.WEIGHTING403.PROBES404.OPTIMIZER405.ALIASING406.DEBUGGING407.DISPATCH408.PIECEWISE409.COMMUT-A TIVE410.CONVOLUTION411.DISKS412.BUGS413.INEQUALITIES414.FOOTPRINT415.TRANSIENT416.SUBCLASS417.BROWSING418.ENCODINGS419.SUPPRESSION420.COMPROMISED 421.PROVABLE422.PARSER423.SEGMENTED424.CONTINUATION425.PERTURBATION426.SLOWDOWN427.HETEROGENEITY428.HARMONIC429.FORMANCE430.VOID431.ADVERSAR IAL432.INJECTED433.COMPACTION434.CPUS435.AGGRESSIVE436.LAYOUTS437.SKELETO-N438.PROP439.SURVEILLANCE440.PRECOMPUTED441.TRANSITIVITY442.ATTACKERS443 .DENSITIES444.FILTERED445.HORN446.FOURIER447.CLIP448.MAXIMA449.TAYLOR450.STRAINTS451.STACKS452.STACKS453.INTENSITIES454.ASSERT455.COMPREHENSION456.AB SORPTION457.MASKS458.SPONDING459.ANOMALOUS460.BENIGN461.MASKING462.FUZZY463.SNAPSHOTS464.SUBSUMED465.MENTATION466.SUBPROBLEM467.FOREACH468.PR OVABLY469.PACKET470.PACKETS471.CLIENTS472.OPTICAL473.WORKLOADS474.SLOT475.CLASSIFIER476.TEXTURE477.GRADIENT478.SUBTREE479.FORWARDING480.GPU481.VID EOS482.REACHABILITY483.CLASSIFIERS484.FOREGROUND 485.ADVERSARY486.CONTENTION487.PEERS488.BACKUP489.TOKENS490.ENCRYPTION491.PROBING492.ATTACKER493.QUANTUM494.NONEMPTY495.DEPLOYMENT496.CALIBRA TION497.TRANSITIVE498.MOBILITY499.VELOCITY500.PORTS501.OVERHEADS502.REPLICAS503.PASSWORD504.EIGENVALUES505.COUNTERS506.SUBTREES507.MODALITY508.NO NDETERMINISTIC509.CACHES510.SMOOTHNESS511.RECEIVERS512.BITMA513.OBLIVIOUS514.HANDLER515.REPLICA516.ENCRYPTED517.CACHING518.GPUS519.CIRCUITS520.TER ING521.LITERALS522.POISSON523.MULTICORE524.MULTISET525.PREDEFINED526.HAMMING527.INJECTIVE528.SUBSPACES529.COMPOSITIONAL530.DISJUNCTION531.SYNCHRON OUS532.LIVENESS533.SUBSYSTEM534.COSINE535.KEYBOARD536.ROBOT537.PROJECTOR538.TIMEOUT539.PHOTOS540.ANALYZER541.THRESHOLDING542.CLIPS543.CONSTRUCT OR544.BOOSTING545.MANUSCRIPT546.LITERAL547.LOOKUPS548.MULTITHREADED549.DISSIMILARITY550.PYRAMID551.CELLULAR552.BAYES553.CENTROID554.SNIPPETS555.PA SCAL556.SELECTIVITY557.GARBAGE558.CLICK559.MULTIPLEXING560.FISHER561.BIPARTITE562.RITHM563.ANCHOR564.CARLO565.OPERANDS566.BUG567.MONTE568.CACHED56 9.HILBERT570.AMPLITUDE571.ZOOM572.ROTATIONS573.BUFFERED574.OBLIGATIONS575.TERNARY576.POLYGON577.SYNTACTICALLY578.FETCH579.ROBOTS580.INRIA581.TIMES

TAMPS582.WEBPAGE583.CATEGORICAL584.LOWING585.OPPONENT586.SUBDIVISION587.COMMUTE588.DURATIONS589.MATLAB590.REWRITES591.DOWNSTREAM592.PREMISES593.HEALTHCARE594.REALISM595.COMPLIANCE596.SPEEDUPS597.MINIMA598.ROTATED599.STA-NFORD601.HALTS602.CUBE603.PRECOMPUTATION604.LOSSY605.OVERLOADED606.POINTWISE607.TRANSLATOR608.HASHES609.STATEFUL610.REPEATABILITY611.BINOMIAL612.COUNTEREXA-MPLE613.ORDIAL614.BIOLOGY615.EFFICACY616.KOLMOGOROV617.LOGISTIC618.DIGITS619.MAGNITUDES620.INVERTIBLE621.PORTABLE622.SOCKET623.AUTOMATON624.SOUNDNESS625.PEER626.MULTICAST627.CONGESTION628.PROVENANCE629.HOP630.REGULARIZATION 631.CONJUNCTIVE632.FINITELY633.DISCRIMINATIVE634.AUTHORIZATION635.CURVATURE
636.EXISTENTIAL637.GUI638.WORKFLOW639.GRADIENTS640.RATINGS641.CLIQUE642.TRACKERS643.LENS
644.PROFILES645.AFFINE646.SCHEMAS647.CONTROLLERS648.WAVELENGTH649.UNSUPERVISED

650.ANGULAR 651.TURING 652.CRYPTOGRAPHIC653.LOGGING654.EIGENVEC-TORS655.AESTHETIC
656.DESTINATIONS657.RENAMING658.LAPLACIAN659.EIGENVALUE660.SPARSITY661.NONBLOCKING662.INDUCTIVELY663.OPERAND664.INFECTE-D665.ISOMORPHI-SM666.UNFOLDING667.BA668.AUTHORIZED669.DILATION670.CONCAVE671.PETRI672.QUANTIZATION673.GHOST674.ONTOLOGY675.IMPAIRED676.UPLOAD677.DERIVABLE678.PROJECTIVE679.MISALIGNMEN-T680.BUFFERING681.DISCRETIZATION682.SEGMENTATIONS683.DEADLOCK684.ANDROID685.CALLER686.INFINITY687.GORITHM688.ALIGNMENTS689.NONDETERMINISM690.NAT691.WARPS692.VORONOI693.LOGIN694.POLYGONS695.MULTISETS696.CODEWORDS697.PIE698.FAKE699.CERTIFICATE700.BROWSER701.CARRIER702.CHORD703.LANDMARKS704.RADIAL705.INJECT706.SLACK707.LEXICAL708.TIMELINE709.TUTORIAL710.CONFIGURED711.QUOT-IENT712.TRANSDUCER713.BIJECTION714.STRIDE715.IMPORTS716.ALIAS717.VORTEX718.DIRECTORIES719.POLYNOMIALLY720PROXIES721.DESCENT722.LINEARIZATION723.TRIANGULATION724.TRIBUTED725.MESSAGING726.MULTIPATH727.PARETO728.EXPRESSIBLE729.BASELINES730.PERPENDICULAR731.TRADEOFFS732.ANIMATIONS733.BLOG734.LAGRANGIAN735.QUANTILE736.STEREO737.SUFFIX738.CURSOR739.TENSOR740.TRACKER741.DECIDABLE 742. PROPOSITIONAL 743. ROUTERS 744..WARPING745.VMS746.OVERFLOW747.VARIATIONAL 748.PREORDER
749.UNDECIDABLE750.FORENSIC751.GATEWAY752.STIMULI753.VIRTUALIZATION754.SUBSCRIPTION755.ALGEBRAS756.PEDESTRIAN757.TRAITS758.PERCEPTUAL759.SCHEDULERS760.BIOMETRIC761.DISJUNCTIVE762.GOSSIP763.LEARNABLE764.CHECKPOINT765.STROKES766.CONSTRUCTORS767.PHYSIOLOGICAL768.HASKELL769.FINGERPRINT

Appendix B (333 words in AWL)

1.GRAPH2.LINEAR3.MATRIX4.MAPPING5.SPATIAL6.STATIC7.METRIC8.TEMPORAL9.PARTITION10.CALCULUS11.ROBUST12.TRANSACTIONS13.SWITCH14.TRANSACTION15.PREDICATE 16PROBABILISTIC17LATENCY18.WORKLOAD19.PREFIX20.CONFIGURATION21.FEASIBLE22.COROLLARY23.MAXIMAL
24.SCALING25.GEOMETRY26.FRAGMENT27.PREDICATES28.PROCESSOR
29.WEIGHTED30.GEOMETRIC31.VERIFY32.ARRAY33.FRACTION34.HEURISTIC35.GENERIC36.HYBRID37.OVERVIEW38.VERIFICATION39.ALIGNMENT40.ILLUMINATION41.SYMMETRIC
42.CONNECTIVITY43.EXPONENTIAL44.FRAGMENTS45.SIMULATIONS46.LATENT47.COGNITIVE48.RENDERING49.MERGE50.DIFFERENTIAL51.TEMPLATE52.ADDITIONALLY53.DEFAULT54.ANNOTATIONS55.ANNOTATION56.AUTHENTICATION57.AUXILIARY58.EQUILIBRIUM59.INVERSE60.TRAJECTORY61.TRAJECTORIES62.CONJUNCTION63.MAPPED64.HORIZONTAL65.ANOMALY66.REDUNDANT67.ANNOTATED68.DEPENDENCE69.REGRESSION70.DISPARITY71.CORRESPONDENCES72.BOTTLENECK73.CONGRUENCE74.DIAGRAM75.PROBING 76.PRIMITIVE77.DIVERGENCE78.INCREMENTAL79.PROBE80.GRAPHICS81.COLLABORATIVE82.GENERATOR83.SPECTRUM84.NAVIGATION85.SEQUENT86.HETEROGENEOUS

87.QUANTITATIVE88.AXIOMS89.LOGICS90.PARALLELISM91.LOCALITY92.MANIFOLD93.SUBJECTIVE94.VIEWPOINT95.SUFFICES96.ALIGNED97.COMPUTATIONALLY98.IDENTIFIER99.SYMMETRY100.INTERPOLATION101.VERIFIED102.SPAN103.CORRUPTION104.NEGLIGIBLE105.POSITIVES106.CONTAINMENT107.TRADEOFF108.CONSERVATIVE

109.STATIONARY110.INSTRUMENTATION111.PROFILES112.ELECTRONIC113.HOMOGENEOUS114.DECODING115.DIFFUSION116.GENERALIZE117.LOCALIZATION118.SYNTHESIS119.SUBSTITUTIONS120.DELETE121.ARTIFACT122.ARRAYS123.EQUIVALENTLY124.FOSTER125.ASSERTION126.ASYMMETRIC127.DESCENDANT128.PROGRAMMER129.REDUNDANCY130.VARIATIONAL131.AFFINITY132.CONVERGE133.TEMPLATES134.EQUALITIES135.CONTROLLERS136.CONTOUR137.INTER-138.CENTRALIZED139.CUMULATIVE140.VULNERABLE141.FUSION142.LEGITIMATE143.PRECONDITION144.TIMING145.INVOCATION146.SIMULATOR147.PRIORITY148.SIMULTANEOUS149.DIRECTORY150.PROXY151.VIRTUALIZATION

152.PROPAGATE153.ESTIMATOR154.DELETIONS155.INITIALIZE156.COLLABORATION157.SALIENT158.BUDGET159.SCALED160.TRANSPARENT161.VOCABULARY162.PREMISE

163.SUPERVISED164.VULNERABILITY165.STATICALLY166.PREDICTOR167.SPATIALLY168.OBJECTIVES169.EMERGENCY170.INCURS171.WEAKER172.PROXIMITY173.INVERTED174.TANGENT175.INVALID176.VULNERABILITIES177.TERMINOLOGY178.REPLICATED179.NEGATION180.ACCELERATION181.ASSERTIONS182.HUGE 183.ACTIVATION 184.DIFFUSE 185.AFOREMENTIONED 186. FOCAL
187.CARDINAL188.SINGULAR189.FORMALIZE190.INCREMENTALLY191.EPOCH192.TOLERANCE193.CAUSAL194.INFORMATIVE195.DUALITY196.COUNTERPART

197.FORMALIZED198.LATENCIES199.ASYMPTOTICALLY200.ENTAILS201.FRAGMENTATION202.DESIGNATED203.QUANTIFICATION204.ACTIVATED205.RESILIENT206.FLUID

207.EMBEDDINGS208.VALUATIONS209.INVOCATIONS210.ADJACENCY211.FEASIBILITY212.INCUR213.PROJECTIVE214.CORRUPTED215.DISSEMINATION216.DEVELOPER

217.GENERATIVE218.PERSONALIZED219.AUTHENTICATED220.FRACTIONAL221.RATIONALE222.ELABORATE223.PREDICTORS224.SELECTIVITY225.INCURRED226.CONTIGUOUS

227.CONVERGENT228.FORMALISM229.PLAUSIBLE230.ABORT231.ENTERPRISE232.ALIGN233.INFEASIBLE234.ADAPTIVITY235.BLURRED236.SIGNALING237.ACCURACIES

238.QUANTIFIED239.VIEWPOINTS240.IDENTIFIERS241.DEGRADE242.INSTANTANEOUS243.CONJECTURE244.VERIFIER245.PROGRAMMABLE246.ASCENDING247.ASCENDING 248.ASCENDING249.OBSTACLE250.OBSTACLES251.TERMED252.CONCISE253.PROMINENT254.PROMINENT255.CHARACTERIZING256.TEMPORALLY257.MATCHINGS258.CAUSALITY 259.ELAPSED260.PRESCRIBED261.RESOLUTIONS262.SPURIOUS263.INHERITANCE264.NUMERIC265.TRANSPARENCY266.IMPERATIVE267.PRECONDITIONS268.SIMPLIFYING 269.RESILIENCE270.TACKLE271.SYNTHESIZED272.AMPLE273.ALARM274.DESCENDANTS275.INHERITED276.LOCALIZED277.DECODE278.FORMALIZATION279.OPTIMALLY280.EQUIVALENCES 281.GEOGRAPHIC282.ENUMERATE283.PARENTHESES284.ASSOCIATIVE285.DEGRADE286.DEVISE287.TOLERATE288.INTRACTABLE289.INCREMENTS290.CREDIBILITY 291.INTERSECTIONS292.AUGMENT293.EXTRAPOLATION294.VERIFIES295.DISTRIBUTION296.RECOGNIZABLE297.TRUNCATED298.DIALOG299.EMULATED300.POSITIONAL 301.REPOSITORY302.BIOLOGICAL303.BIOLOGICAL304.IMPORTS305.COMPILERS306.QUANTITATIVELY307.VARIANCES308.ANTECEDENT309.CONSOLIDATION 310.PROXIES311.AUTHENTICATE312.READABILITY313.REPLICATE314.MODULATION315.AMENABLE316.ADAPTIVELY317.CONDITION318.CONSERVATIVELY319.AUGMENTING320.

ITERATE321.ARTICULATED322.HOLISTIC323.ROBUSTLY324.SIMPLIFIES325.SPONTANEOUS326.STANCE327.CONJUNCTIONS328.HEREAFTER329.SKEWED330.SUBTRACTION331.PORTABLE332.REPETITION 333.TAXONOMY

Appendix C (50 new GSL words)

1.PROFILE 2.NEIGHBORING 3.ENCODE 4.PLOT 5.PLATFORM 6.RANDOMIZED

7.INSPIRED 8.OPT 9.FREQUENCIES 10.CAST 11.SOPHISTICATED 12.MOVIE

13.POTENTIALS 14.BUDGET 15.EMAIL 16.RENDERED 17.WEIGHTING

18.PHOTO 19.EXPRESSIVENESS 20.INTERESTINGLY 21.PHRASES 22.PATIENTS

23.HUGE 24.RENDER 25.PHRASE 26.CHIP 27.WEAKENING 28.GENUINE 29.BARRIERS

30.CLICK31.MISSION32.APPEAL33.ENGAGE34.GUIDANCE35.VAST36.BINDINGS

37.APPEALING 38.REFERENCED 39.COMBINATORIAL 40.DISABLED 41.BRAND

42.ENGAGED 43.ABUSE 44.RECORDINGS45.PROTEIN46.BOOST 47.STALE 48.WEAKEST

49.REPETITIVE 50.VICTIM