

A CORPUS ANALYSIS OF GRANT GUIDELINES: THE EDUCATION AND TRAINING PROGRAMME WORD LIST (ETPWL)

Zsófia K. Freund

Department of English Applied Linguistics, Faculty of Humanities, University of Pecs, Hungary
Phone: +36305179717, E-Mail: zsofiafreund@gmail.com

Abstract. *This study created an education and training programme word list (ETPWL) to explore the characteristics of lexical elements of the official document of the latest calls for proposals of Education and Training Programme (ETP) proposed by the European Commission. The aim of this investigation is to support the assumption that these official documents have EU-specific characteristics, what is more, grant-specific features. The researcher compiled the Corpus of Education and Training Programme (CETP), containing 5 sub-grant calls in 2013, (totaling 252,599 running words, over 5,891 word-types) and examined the range, frequency and usage patterns of lexical elements in CETP and the British National Corpus (BNC). The comparison is of usage patterns in CETP and BNC written uncovered lexical patterns that are typically used in relation with calls for proposals and grant applications. The vocabulary of CETP is not specific in terms that it does not require the knowledge of specific terminology which would not be known to non-native speakers with advanced level of English knowledge, but the high frequency number shows that these lexical items need special attention by the applicants. This study identifies 604 words that are believed to be indispensable for those who are going to apply for such grants in the future.*

Key words: *English for Specific Purposes, corpus-based research, word list, EU English*

1. INTRODUCTION

The rise of the use of English within the European Union (EU) has brought the need for EU specific English. The English language has become a lingua franca among the EU member states. It has not only affected the general public and the business world, but also the status of English as an integral part of professional life (McArthur 2003; Modiano 2001; Seidlhofer 2001; Seidlhofer 2004; Seidlhofer, Breiteneder and Pitzl 2006.)

In the last decade language use within the EU has been in the centre of research among applied linguists, terminologists, translators and among those who are working with ESP. Until now, only a few studies have investigated the language of grant proposals (Connor and Mauranen 1999; Connor 2000; Feng and Shi 2004; Tseng 2011). The linguistic features of grant proposals were investigated by Connor and Mauranen (1999) to extend academic and researcher awareness about the significance of persuasive writing in grant proposals.

Grant guidelines are important academic genres in many disciplinary areas. No doubt that grant proposals have become a significant part of professional writing; academics in all fields sooner or later encounter grant guidelines (Connor and Mauranen 1999; Connor

2000). These official documents have their own specificities in terms of the use of EU English. The term is used by Trebits (2008, 2009a, 2009b) and Jablonkai (2009, 2010) in order to distinguish general English and English which is used in EU documents. As English has become the lingua franca in Europe, English teachers need to prepare for the linguistic challenges the learners may face. There have only been a few studies on investigating EU English used in documents for pedagogic purposes (Jablonkai, 2009, 2010; Trebits, 2008, 2009a, 2009b), but no research has been made to analyse the usage patterns in grant calls.

The present study aims to compile a word list which contains lexis that future applicants need to be familiar with in the application process. A specialized corpus, named Corpus of Education and Training Programme (CETP) has been compiled to identify the most frequent lexical items and to make it possible to compare them to the General Service List (GSL)/Academic World List/(AWL) and BNC lists with the help of a corpus analysis software. Lexical items with high number of occurrences have been compared to their frequencies in the British National Corpus (BNC Written) to reveal how word usage of CETP differs from a general English text. The findings of this investigation can be useful for designing course materials for English EU grant writing courses and is considered to be indispensable for those non-native speakers of English, wishing to submit a project proposal successfully.

2. COMPILING AN EDUCATION AND TRAINING PROGRAMMES WORD LIST (ETPWL)

2.1. Research Aims

The investigations intend to uncover the most frequent lexical elements and present the analysis of their usage patterns by establishing a specific word list of grant-related vocabulary, based on grant application guidelines and calls published by the Education and Training Programmes (ETP) of the European Commission (EC). The Corpus of Education and Training Programmes (CETP) helps identify the most frequent content words and presents a possible analysis of some of their usage patterns in grant calls.

The basic idea was to look into differences between the vocabulary used in grant calls published by the EC under the ETP and the vocabulary of general English texts from the database of the BNC. The hypothesis was that the official documents of grant applications and calls used specialized language that non-native speakers must be familiar with in order to interpret the text adequately.

The investigation focused specifically on the following issues:

1. What are the most frequently used lexical elements in the Corpus of Education and Training programmes (CETP) and what differences are noticeable in comparison to the reference corpus?
2. What are the most frequently used lexical elements in the CETP which are not among the first 2,000 most frequent words of English as given in the GSL?
3. What are the most prevailing usage patterns of frequent lexical items in the CETP?
4. Are there any grant-specific lexical elements in CETP and how frequent are these?
5. What are the pedagogical implications of the findings for future ESP courses?

2.2. Creating the Corpus of Education and Training Programme (CETP)

Every year, the European Commission sets the priorities for the EU's ETP in calls for proposals. The Programme's major aim is to give funding to improve the quality of the educational system. The grant guidelines we focused on concern higher education primarily. These co-ordinated actions are designed to closely work with national authorities on improving policies with partner institutions from other countries (Education & Training Programme, n.d.). The five sub-programmes and actions promoted were chosen under the Education and Training Programme: Lifelong Learning, Erasmus Mundus, Tempus, Bilateral cooperation and Intra-ACP academic mobility scheme.

The documents related to the latest calls in all sub-programmes under the Education and Training Programme were collected for typifying a specialized language constituents which this paper aims to identify. The data in the present study came from the official webpage of the European Commission. The compiled specialized corpus contains 252,599 running words. It is widely believed that a large number of samples are needed to draw statistically valid conclusions; however, the size in itself is no guarantee for reliability. The compiled data represents the characteristics of the language of the calls and guidelines of ETP of the EC. Therefore, the findings from the analysis are believed to be representative and to be generalizable to the previous calls as well. The analysing software was Compleat Text Stripper.

3. RESULTS

3.1. Counting words/words that count

The Corpus of Education and Training Programme (hereafter referred to as the CETP) was specifically compiled for this study. The text of guidelines formed a corpus of 252,599 tokens and over 5,891 word-types. Frequently occurring words, being a useful starting point in identifying a unique or shared vocabulary in the texts were investigated. Not only content words can reflect the specificity of a particular text or field, but there might be grammatical words which carry more subtle information. In the case of specialized corpora, a higher proportion of content words are expected. A comparison between grant guideline texts and general English texts reveals the most important usage patterns of those lexical items which are frequently found in the CETP but not frequent in the BNC.

A word frequency analysis of the CETP was carried out in comparison with the GSL, the AWL and the BNC 14,000 lists from Nation's RANGE software. In order to present the percentage of the individual coverage of the 10 most frequent words, a frequency word list was created with Nation's Frequency program. The 10 most frequent lexical items have been chosen for comparison with Jablonkai's (2009) findings (Table 1), in which she collected the most frequent words in EU discourse. The analyses of the created frequency list have revealed that the most recurrent words are function words. The comparison has also showed that the first seven items in both frequency lists are the same. Furthermore, two items are shared ones among the remaining three items, but with different frequency ranks. This is in line with previous expectations and results: in most corpora these function words occupy the first few places on frequency lists (Jablonkai, 2009; 2010; Rizzo, 2010).

Table 1 The 10 most frequent words in CETP and Jablonkai's corpus (2009, p. 8)

Most frequent words in CETP			Most frequent words in EU discourse				
RANK	FREQ	%	WORD	RANK	FREQ	%	WORD
1.	20435	8.09	THE	1.	10142	8.50	THE
2.	11363	4.50	OF	2.	5589	4.68	OF
3.	9821	3.89	AND	3.	4071	3.41	AND
4.	6172	2.44	TO	4.	3184	2.67	TO
5.	5890	2.33	IN	5.	2684	2.25	IN
6.	4692	1.86	FOR	6.	2023	1.70	FOR
7.	3572	1.41	A	7.	2001	1.68	A
8.	2927	1.16	BE	8.	1355	1.14	ON
9.	2157	0.85	OR	9.	1075	0.90	BE
10.	1822	0.72	ON	10.	957	0.80	BY

As shown by Table 1, THE is the most frequent word in the corpus and stands for more than 8% of the total tokens in both corpora. The following OF and AND function words have similar percentage figures in both registers.

The GLS/AWL Range programme provides information about what percentage of the items in the text is covered by three built-in lists. 'One' means the first 1,000 words, 'two' gives the second 1,000 words of the GSL and 'three' indicates the ratio of the academic words from the AWL. The 'not in the lists' or 'off-list' category gives the ratio of rare words, i.e. words not included in any of the previous three lists. Investigation uncovers that 170,401 of the running words in the text are in base list one and make up 67.46 percent of the tokens in the CETP. Surprisingly 469 word families out of the 570 AWL word families appeared in the CETP. This indicates that CETP mostly contains academic vocabulary. As far as the word types are concerned, less than 50 percent of the word types belong to the first 2,000 word types of the GSL. The same analysis was carried out with Nation's RANGE program with the British National Corpus 14,000 list. The result pointed out that most of the frequently occurring words in CETP came from the first two base-lists of BNC 14,000 list. The comparison of the BNC 14,000 and the CETP showed that the total BNC 14,000 contains approximately 60 percent of the word types in the CETP. It also indicates that only 6.74 percent of the word types are academic and the highest proportion, 39.47 percent are off-list words. Rare words give more than one-third of the total coverage. This means that a high proportion of the words in CETP are specialized.

Nation's Frequency program makes it possible to draw up a list of frequently occurring words in alphabetical order or creates a frequency-ordered list. A frequency-ordered word list of the CETP displays the word types and families the corpus contains and ranks them according to the number of occurrences of each word type and word family. The words included in the GSL first and second base lists were eliminated and only word types and families were used, which have been found in base list three and types which have not been found in any lists.

It is not surprising that the most frequent content words are PROJECT and GRANT. The most frequent word families include PARTNER, CRITERIA, DURATION, RELEVANT, SPECIFIC, MINIMUM, PERIOD, RESEARCH, ADULT and APPROPRIATE. A semantic analysis performed on this section of the corpus resulted in identifying five major semantic sets: EU funding, legal, integration, communities and other. The word families indicate that the listed lexical items are mostly legal words occurring in a wide range of EU-related subject fields such as AUTHORITY, PRINCIPLE, PRIORITY, POLICY, CONSULT, LEGAL, SECTION, etc. The high text coverage of legal lexis can be attributed to the fact that these guidelines are based on formal legalisation texts (Council decisions) and their function is to guide the applicant how to follow a certain rule or policy in the application process. The semantic analysis also reveals that there are various, not surprisingly, EU funding related words like PROJECT, GRANT, MONITORING, RESEARCH, OBJECTIVE and TIMETABLE; and vocabulary expressing aspects of integration such as COOPERATION, COORDINATION, PROCESS and PARTNER. The following words refer to the communities within the EU: TARGET, NETWORK, PARTICIPATE, COMMISSION.

The same analysis was performed by the RANGE BNC 14,000 list. As far as the semantic analysis is concerned, a few examples can be found for EU funding vocabulary: BUDGET, ELIGIBLE, AWARD, some words referring to integration are ADAPT, GLOBAL, ABROAD, ORIGIN and HOST, while COORDINATE, SUPERVISE and FRAMEWORK belong to the semantic set of communities. Finally, PENALTY, TERMINATE and CORRESPOND are legal words.

The words not found in any list mostly involve geographical names (WASHINGTON, VIENNA) and abbreviations/acronyms, for example LLP, ETC, SME, EC, RD and EUR, but these were omitted from the lists. Interestingly, a high proportion of off-list words are the same such as EU, ELIGIBLE, DISSEMINATION, MULTILATERAL and CONSORTIUM. These words are in connection with applications for EU funding. Some differences are recognizable, such as the word BUDGET, which appeared in the rare word list in the GSL/AWL list, but in the BNC 14,000 list, RANGE placed BUDGET into the first 1,000 list.

3.2. Findings on patterns of use

After examining the distribution of academic and off-lists word lists the usage patterns of ten lexical items from both the CETP and BNC Written were compared. The aim of this analysis was to identify specific lexical patterns characterizing written ETP discourse. The assumption was that even if CETP shares words with BNC Written, their usage patterns or meanings may differ. The present study focused on ten lexical items selected based on the following criteria:

- Selected items had to be words with high-frequency and excluded from GSL representing the first 2,000 most frequent English words
- A minimum of 15 occurrences, both in CETP and BNC written.

The Concordancer program of Compleat Lexical Tutor was used for analysing discourse functions of the selected items in the two registers, as the concordance list can display words with their surrounding context. Table 2 shows which lexical items have been compared by the Text Based Concordancer and Corpus Based Concordancer programs. As far as the meaning of the words are concerned, the official website of the European Union ([www.europa.eu](http://europa.eu)) provides Euro-Jargon (http://europa.eu/abc/eurojargon/index_en.htm)

and EU glossary (http://europa.eu/legislation_summaries/glossary/) sections for the basic and essential terms used in the official documents. The Cambridge Online Dictionary (hereafter referred to as the COD) (available at <http://dictionary.cambridge.org/>) was also used in defining the meaning of certain words.

Table 2 Selected lexical items for analyses of their usage patterns

	Word Family	Number of occurrences of word families in CETP	Number of occurrences of word families in BNC Written	Overrepresentation in % (CETP/BNC Written)
1.	ELIGIBLE	942	15	6280%
2.	CRITERIA	527	23	2291%
3.	COOPERATE	390	28	1393%
4.	PRIORITY	393	48	819%
5.	OBJECTIVE	447	63	710%
6.	IMPLEMENT	318	55	578%
7.	SUSTAIN	163	37	441%
8.	ENHANCE	114	33	345%
9.	CAPACITY	149	45	331%
10.	AMEND	28	30	93%

A close examination of the concordance lists allowed for sorting out how different or similar the meanings of the same lexical elements were in the CETP corpus. From the examination of different usage patterns and word types in the two corpora, the following observation has been made: there are minor differences in terms of meaning, but the main differences appeared in the number of occurrences which indicates that some words occur more often in specialised text than in everyday language.

In terms of usage patterns, a following observation has been made: unlike the general usage of the lexis OBJECTIVE, in grant guidelines it rarely appears in itself but collocates with various adjectives referring to the various level of the realization of the goals of the project (for example general, overall or specific objectives), and the following “of” possessive is also common. In BNC Written CAPACITY is used mostly in its first meaning (amount that can be contained) in phrases like “storage CAPACITY”, “spare CAPACITY”, “runway CAPACITY” and “nuclear CAPACITY”, etc. On the contrary, in CETP the word CAPACITY is only used in a figurative sense, when describing the necessary endowments one has to have to create a project, e.g. “management CAPACITY”, “financial CAPACITY”, “administrative CAPACITY”, “operational CAPACITY” and “professional CAPACITY”. The word family of SUSTAIN in BNC is used mostly in an economic sense such as “SUSTAINABLE growth”, “SUSTAINABLE economy”, or “SUSTAIN budget deficit”, while in CETP the word family, though its primary meaning is intact, refers to the sustainability of project results and indicators after the lifetime of the project, therefore its meaning is more specific than in BNC. Seeing the hits for the word IMPLEMENT in CETP, two remarks must be made: the word is either used as IMPLEMENTATION (as a noun) or be IMPLEMENTED, in the passive voice. In BNC there is no occurrence to see this word family being used in its active form (x is IMPLEMENTING y). As much as the context is concerned, the word is used when referring to carrying out something according

to rules set externally. Similar to the word SUSTAIN, PRIORITY is more a grant-related word. In BNC it is usually preceded by adverbs like “high” and “immediate” or used in the passive voice. In CETP its usage pattern is slightly different, it is used to start definitive statements, such as “the PRIORITIES are...”. This word is generally used to describe the main status to achieve in order of importance or urgency through project calls. Regarding the words ELIGIBLE, CRITERIA, ENHANCE, AMEND, COOPERATE, there is no real discrepancy in terms of patterns of use and it can be seen that the same idea is intended.

3.3. Establishing the Education and Training Programme word list (ETPWL)

Coxhead’s (2000) three selection criteria were adopted with some modification for the compilation of ETPWL in this study. The first criterion indicates that in order to identify the specific vocabulary in the guidelines of ETP, those content words have to be identified first which are not among the first 2,000 most frequent words of general English as given in the GSL. The second criterion, range, would record the number of files in which each word occurs. In the case of ETPWL development this criterion is not relevant because CETP is only one file. Finally, the frequency criterion was established, as a word family member had to occur at least ten times in the CETP to be selected into the word list.

The final selection criteria were formulated as follows:

1. Specialised occurrence: word families included in the final ETPWL had to be beyond the GSL representing the first 2,000 most frequent English words and beyond the first 2,000 English words in the BNC 14,000 list.
2. Frequency: members of a word family have to occur at least 10 times in the corpus.

After the words were carefully selected, their family members were chosen, but plural forms were omitted in the compilation of the ETPWL. Bauer and Nation (1993) argue that a word family includes a base word and its derived and inflected forms. This indicates that it does not require extra effort from a language learner to understand the derived or inflected word forms as they are familiar with the base word. The following words have been omitted, because these were not considered relevant enough in the ETPWL, such as EU, ADULT, WEB, INTERNET, ANNEX and CHAPTER. Geographical names, acronyms, abbreviations and sub-program names were also excluded from the list. The resulting list was termed as the Education and Training Programme Word List (ETPWL) and it contains 604 words. The list is attached in Appendix A.

3.4. Discussion of findings

The first research question investigated which lexical items in CETP had the highest frequency rank. Table 1 illustrates that the first ten words on the list are function words. The findings support results of earlier studies (Jablonkai 2009; Jablonkai 2010; Rizzo 2010), where the examination showed that the first 100 words are functional in a general corpus, but it is also expected that the more specialized a corpus is, the more content words can reach high frequency occurrences. The present investigation also concluded that the first fourteen words in the frequency lists were function words.

Nation’s Lexical Frequency and RANGE software have uncovered that almost half of the lexical elements in CETP were not in the GSL or among the first 2,000 words of the BNC list. Surprisingly, 469 word families out of the 570 AWL word families were used in the official documents of calls for proposals. The investigation showed that 20 percent of

word types in the CETP are academic. These findings are very similar to those of Wang et al. (2008) and Vongumivitch et al. (2009) on the comparison of the Medical Academic Word List and the Academic Word List, and between the Applied Linguistic Corpus and the Academic Word List. Mudraya's (2006) corpus also proved that many overlaps could be seen between the Student Engineering English Corpus (SEEC) and Coxhead's AWL (2000). Mudraya (2006) pointed out that more attention should be devoted to the academic words listed in the AWL. Vongumivitch et al. (2009) showed an 11.17 percent overlaps between the AWL and Applied Linguistics Research Articles Corpus (ALC). Based on their literature review and their study, Vongumivitch et al. (2009) concluded that the "AWL plays an important role in specialized texts" (p. 35). This statement is supported by the present investigation too, which shows that non-native speakers of English, with general academic knowledge, would be likely to understand the vocabulary of grant calls and guidelines. A few non-AWL content words, which occurred with high a frequency, were also considered in the compilation of the word list. More than 40 percent of the word types in CETP occurred in the off-lists in both reference corpora. A closer investigation showed that many of the off-list words were abbreviations, geographical names or names of the sub-programmes of the Education and Training Programmes.

According to Flowerdew (2005), in order to identify a specialized vocabulary it is not enough to look at the frequency list of a specialized corpus, but it needs to be compared to reference corpus. In his investigation Flowerdew (2005) concluded that the difference between a specialized corpus and a general corpus might be in the usage patterns of certain lexical items. The present study compares CETP with BNC Written to see whether it can support Flowerdew's findings, too (2005). The concordance lists from the BNC Written also proved to be helpful in exploring interesting word usage patterns. Hits for ten word families were examined to compare some of the most frequently occurring words in both corpora. This investigation aimed to illustrate how the use of lexical elements in particular registers could differ. The comparisons showed that the majority of the occurrences of the ten word families behave in roughly similar ways in terms of meaning across the two registers in the case of the word families ELIGIBLE, CRITERIA, AMEND, COOPERATE, ENHANCE. The usage patterns of the word families OBJECTIVE, SUSTAIN, CAPACITY, IMPLEMENT and PRIORITY are slightly differed in the two registers.

Hyland and Tse (2007) pointed out that "all disciplines adapt words to their own ends, displaying considerable creativity in both shaping words and combining them with others to convey specific, theory-laden meanings associated with disciplinary models and concepts" (pp. 245-246). This argument proved to be true in this case as well. A closer examination of the lists showed that many words in the CETP were considered as EU and grant specific. The analysed words occurred far more often in CETP than in general English discourse. These words do not belong to the general English vocabulary of an average learner of English because these lexical items occurred outside from the first 2,000 words of the GSL. Knowing the 2,000 words of the GSL, it would be difficult to interpret the ETP guidelines, but with an academic vocabulary it should not cause difficulties. Eight out of the ten examined words belong to the AWL. The vocabulary of CETP does not require the knowledge of highly specific terms, however the high frequency number shows that these lexical items need special attention by the applicants and their mastery is essential for those who wish to apply for such a grant. The present paper has provided a list of 604 words (ETPWL) that are more frequently used in grant guidelines of ETP than in general English texts.

Joining the EU brought new grant opportunities; therefore knowing EU English or being familiar with grant specific vocabulary is essential. Krishnamurthy and Kosem (2007)

emphasized that syllabus and course material design, language testing and classroom methodology are the areas that have great potential in corpus research from the pedagogical aspect. The compiled ETPWL can serve as reference for EU English lexical syllabus design as words in ETPWL are worthy paying special attention to when designing English for Specific Purposes courses. It goes without saying that a thorough needs analysis is a demand. ETPWL also can help applicants acquire specific vocabulary in more conscious and manageable ways in order to adequately interpret calls and guidelines of Education and Training Programmes.

5. CONCLUSION AND FUTURE STUDY

The ETPWL provides a fairly high coverage of calls and guidelines of grants proposed by the European Commission under the Education and Training Programmes. This list can serve as a basis for courses of English for EU grant specific purposes. The research identified 604 words which are believed to be indispensable in text comprehension. This study highlights the importance of developing specialized word lists in ESP for language teaching purposes for different disciplines. The findings of the study may serve as a starting point in developing a corpus from EU grant guidelines.

In the light of the findings of the present paper, many issues need to be taken into consideration in further studies. An obvious limitation of the study is that it deals only with documents of ETP which have limited sources; a larger corpus would allow for making more general observation on all types of grant calls proposed by the European Commission. Nevertheless, this has revealed important characteristics of the official documents of calls of ETP. Although the focus of the present paper was the vocabulary and its patterns of use in grant guidelines of ETP, it is believed that further investigations into other research paradigms such as lexical bundles, conjunctions or syntactic features should be encouraged. The examination of lexical choices in guidelines of grant calls can be extended to investigating grammatical, pragmatic and discourse aspects of self-compiled corpora.

REFERENCES

- Anna Trebits, "Conjunctive cohesion in English language EU documents: A corpus-based analysis and its implications", *English for Specific Purposes*, 28 (3) (2009a): 199–210.
- Anna Trebits, "English lexis in the documents of the European Union: A corpus-based exploratory study", *Working Papers in Language Pedagogy*, 2, (2008): 38–54, accessed November 24, 2012, <http://langped.elte.hu/WoPaLParticles/W2Trebits.pdf>
- Anna Trebits, "The most frequent phrasal verbs in English language EU documents: A corpus-based analysis and its implications", *System*, 37 (3) (2009b): 470–481.
- Averil Coxhead, "A new academic word list." *TESOL Quarterly*, 34 (2) (2000): 213–238.
- Barbara Seidlhofer, "Euro-English: Towards making 'Euro-English' a linguistic reality", *English Today*, 68 (17) (2001): 14–16, accessed November 25, 2012, <https://mapef.ffri.hr/datoteke/Sirola/ee-jenmodseid.pdf>
- Barbara Seidlhofer, "Research perspectives on teaching English as a lingua franca." *Annual Review of Applied Linguistics*, 24, (2004): 209–239, accessed December 18, 2012, http://people.ufpr.br/~clarissa/pdfs/ELFperspectives_Seidlhofer2004.pdf
- Barbara Seidlhofer, Angelika Breiteneder and Marue Liouse Pitzl, "English as a lingua franca in Europe: challenges for applied linguistics", *Annual Review of Applied Linguistics*, 26, (2006): 3–34.
- British National Corpus. Corpus description, accessed March, 2013, http://www.lextutor.ca/concordancers/corpus_descriptions.html
- Camino Rea Rizzo, "Getting on with corpus compilation: from theory to practice", *ESP World*, 1 (27) (2010): 1–23, accessed December 29, 2012, http://www.esp-world.info/articles_27/camino%20rea.pdf

- COD. Cambridge Online Dictionary, accessed March, 2013, <http://dictionary.cambridge.org/>
- Education and Training Programme, accessed March, 2013, http://eacea.ec.europa.eu/education/index_en.php
- European Commission. Official webpage of the European Commission, accessed March, 2013, http://ec.europa.eu/index_en.htm
- Gaetanelle Gilquin, Sylviane Granger, S. and Magali Paquot, "Learner corpora: The missing link in EAP pedagogy", *Journal of English for Academic Purposes*, 6, (2007): 319-335.
- Haiying Feng and Ling Shi, "Genre analysis of research grant proposals", *LSP & Professional Communication*, 4 (1), (2004): 8-30, accessed December 18, 2012, cjas.dk/index.php/LSP/article/download/2013/2013
- Jing Wang, Shao-Ian Liang and Guang-chun Ge, "Establishment of a medical academic word list", *English for Specific Purposes*, 27 (4) (2008): 442-458.
- Ken Hyland and Polly Tse, "Is there an "Academic Vocabulary"?", *TESOL Quarterly* 41 (2) (2007): 235-253.
- Lauri Bauer and Paul Nation, "Word families", *International Journal of Lexicography*, 6 (4) (1993): 253-279, accessed December 8 2012, <https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/1993-Bauer-Word-families.pdf>
- Lynne Flowerdew, "An integration of corpus-based and genre-analysis approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies", *English for Specific Purposes*, 24, (2005): 321-332.
- Marko Modiano, "Euro-English: A new variety of English", *English Today*, 68 (17) (2001): 13-14, accessed November 25, 2012, <https://mapef.ffri.hr/datoteke/Sirola/ee-jenmodseid.pdf>
- Ming-Yu Tseng, "The genre of research proposals: towards a cognitive-pragmatic analysis", *Journal of Pragmatics*, 43 (2011): 2254-2268.
- Olga Mudraya, "Engineering English: A lexical frequency instructional model." *English for Specific Purposes*, 25, (2006): 235-356.
- Paul Nation, RANGE and Frequency Programs, accessed March, 2013, <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Ramesh Krishnamurthy and Iztok Kosem, "Issues in creating a corpus for EAP pedagogy and research." *Journal of English for Academic Purposes*, 6, (2007): 356-373.
- Réka Jablonkai, "English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents", *English for Specific Purposes*, 29 (2), (2010): 253-267.
- Réka Jablonkai, "In the light of: A corpus-based analysis of two EU-related registers." *Working Papers in Language Pedagogy*, 3, (2009): 1-27, accessed November 24, 2012, <http://langped.elte.hu/WoPaLParticles/W3Jablonkai.pdf>
- Tom McArthur, "World English, Euro-English, Nordic English?", *English Today*, 73 (19) (2003): 54-58, accessed November 28, 2012, <https://mapef.ffri.hr/datoteke/Sirola/ee-mcarthur.pdf>
- Ulla Connor and Anna Mauranen, "Linguistic Analysis of Grant Proposals: European Union Research Grants", *English for Specific Purposes*, 18 (1) (1999): 47-62.
- Ulla Connor, "Variations in rhetorical moves in grant proposals of US humanists and scientists", *Text*, 20 (2000): 1-28, accessed December 18, 2012 https://scholarworks.iupui.edu/bitstream/handle/1805/2663/Connor_Variation_Rhetorical.pdf?sequence=1
- Viphavee Vongpumivitch, Ju-yu Huang and Yu-Chia Chang, "Frequency Analysis of the Words in Coxhead's Academic Word List (AWL) and Non-AWL Content Words in Applied Linguistics Research Papers", *English for Specific Purposes*, 28 (1) (2009): 33-41.

APPENDIX A
The Education and Training Programme Word List (ETPWL)

1. ABROAD	50. ATTAIN	99. COMMUNITY
2. ACADEMIC	51. ATTAINMENT	100. COMPATIBLE
3. ACADEMICS	52. AUDIOVISUAL	101. COMPENDIA
4. ACADEMY	53. AUDIT	102. COMPETENCE
5. ACCESS	54. AUDITOR	103. COMPETENT
6. ACCESSION	55. AUTHORISE	104. COMPETITIVE
7. ACCOMMODATE	56. AUTHORISED	105. COMPETITIVENESS
8. ACCOMPANIED	57. AUTHORISING	106. COMPLEMENT
9. ACCOMPANY	58. AUTHORITY	107. COMPLEMENTARY
10. ACCOMPANYING	59. AUTOMATE	108. COMPLIANCE
11. ACCORDANCE	60. AUTONOMY	109. COMPLY
12. ACCOUNTABILITY	61. AVAILABLE	110. COMPONENT
13. ACCREDITATION	62. AWARD	111. COMPOSED
14. ACHIEVE	63. AWARDED	112. COMPOSITION
15. ACKNOWLEDGE	64. AWARDING	113. COMPREHENSIVE
16. ACQUIRE	65. AWARE	114. COMPRISE
17. ACQUIRED	66. BENEFIT	115. COMPULSORY
18. ACQUISITION	67. BILATERAL	116. COMPUTE
19. ADAPT	68. BILINGUAL	117. CONCEPT
20. ADEQUATE	69. BOOST	118. CONCLUDE
21. ADMINISTRATE	70. BRAND	119. CONCRETE
22. ADMINISTRATION	71. BUDGET	120. CONDUCT
23. ADMINISTRATIVE	72. BUDGETARY	121. CONFER
24. ADMINISTRATORS	73. CAMPAIGNS	122. CONFIRM
25. AGENDA	74. CANDIDATE	123. CONFLICT
26. AGRICULTURE	75. CAPACITY	124. CONJUNCTION
27. ALLIANCES	76. CAREER	125. CONSECUTIVE
28. ALLOCATE	77. CATEGORY	126. CONSIST
29. ALTERNATIVE	78. CENTRALISED	127. CONSORTIA
30. AMEND	79. CERTIFICATE	128. CONSORTIUM
31. AMENDED	80. CHALLENGE	129. CONSTITUTE
32. ANALYSE	81. CHAMBER	130. CONSULT
33. ANNEXED	82. CHAMBERS	131. CONTACT
34. ANNOUNCED	83. CHAPTER	132. CONTEXT
35. ANNUAL	84. CHARTER	133. CONTRACT
36. ANTICIPATE	85. CIRCUMSTANCE	134. CONTRACTUAL
37. ANTICIPATED	86. CITIZENSHIP	135. CONTRIBUTE
38. APPROACH	87. CIVIL	136. COOPERATE
39. APPROPRIATE	88. CLARIFY	137. COOPERATION
40. ARCHITECTURE	89. COACHING	138. COORDINATE
41. AREA	90. CODE	139. CORRESPOND
42. ARISING	91. COHERENCE	140. COUNSELLING
43. ASPECT	92. COHERENT	141. COUNSELLORS
44. ASSESS	93. COHESION	142. COURIER
45. ASSIGN	94. COLLABORATION	143. CREATE
46. ASSIST	95. COMMENT	144. CREDIT
47. ASSISTANTSHIP	96. COMMISSION	145. CRITERIA
48. ASSURE	97. COMMIT	146. CRUCIAL
49. ATTACH	98. COMMUNICATE	147. CULTURE

- | | | |
|--------------------|-----------------------|---------------------|
| 148. CURRICULA | 200. ENROLLED | 252. FOSTER |
| 149. CURRICULAR | 201. ENSURE | 253. FOSTERING |
| 150. CURRICULUM | 202. ENTERPRISE | 254. FOUNDATION |
| 151. CYCLE | 203. ENTERPRISES | 255. FRAMEWORK |
| 152. DATA | 204. ENTITIES | 256. FULFIL |
| 153. DATABASE | 205. ENTITLED | 257. FULFILLED |
| 154. DEADLINE | 206. ENTITY | 258. FUNCTION |
| 155. DEADLINES | 207. ENTREPRENEURIAL | 259. FUND |
| 156. DEBATE | 208. ENTREPRENEURSHIP | 260. FURTHERMORE |
| 157. DECENTRALISED | 209. ENVIRONMENT | 261. GENDER |
| 158. DEDICATED | 210. ENVISAGED | 262. GENERATE |
| 159. DEEMED | 211. EQUIP | 263. GLOBAL |
| 160. DEFINE | 212. EQUIVALENT | 264. GLOBE |
| 161. DELEGATIONS | 213. ESTABLISH | 265. GLOSSARY |
| 162. DELIVERABLE | 214. ESTIMATE | 266. GOAL |
| 163. DELIVERABLES | 215. ETHNIC | 267. GOVERNANCE |
| 164. DEMOCRACY | 216. ETWINNING | 268. GRADUATE |
| 165. DEMONSTRATE | 217. EURATOM | 269. GRANT |
| 166. DEPARTURE | 218. EVALUATE | 270. GRANTEES |
| 167. DESIGN | 219. EVALUATION | 271. GUARANTEE |
| 168. DESIGNATED | 220. EVIDENT | 272. GUIDANCE |
| 169. DESIGNATION | 221. EXCEED | 273. GUIDELINE |
| 170. DESTINATION | 222. EXCEEDING | 274. HOST |
| 171. DIALOGUE | 223. EXCLUDE | 275. HUMANITIES |
| 172. DIGITAL | 224. EXECUTION | 276. IDENTICAL |
| 173. DIMENSION | 225. EXECUTIVE | 277. IDENTIFY |
| 174. DIPLOMA | 226. EXPENDITURE | 278. IMPACT |
| 175. DISCIPLINE | 227. EXPERT | 279. IMPLEMENT |
| 176. DISSEMINATE | 228. EXPERTISE | 280. IMPLEMENTATION |
| 177. DISSEMINATING | 229. EXPLICIT | 281. IMPLEMENTED |
| 178. DISSEMINATION | 230. EXPLOIT | 282. IMPLEMENTING |
| 179. DISTINCT | 231. EXPLOITATION | 283. IMPLY |
| 180. DISTRIBUTE | 232. EXTERNAL | 284. IMPORTANCE |
| 181. DIVERSE | 233. FACILITATE | 285. IMPOSE |
| 182. DIVERSITY | 234. FACILITATING | 286. INCENTIVE |
| 183. DOCTORATE | 235. FACTOR | 287. INCOMING |
| 184. DOCUMENT | 236. FACULTIES | 288. INCORPORATE |
| 185. DURATION | 237. FACULTY | 289. INCURRED |
| 186. ECONOMY | 238. FEASIBLE | 290. INDEX |
| 187. EDIT | 239. FEDERAL | 291. INDICATE |
| 188. EDUCATION | 240. FEDERATION | 292. INDIVIDUAL |
| 189. ELECTRONIC | 241. FEE | 293. INDUSTRIALISED |
| 190. ELEMENT | 242. FEEDBACK | 294. INELIGIBLE |
| 191. ELIGIBILITY | 243. FELLOWSHIP | 295. INFORMAL |
| 192. ELIGIBLE | 244. FINAL | 296. INITIAL |
| 193. EMPHASIS | 245. FINANCE | 297. INITIATE |
| 194. EMPLOYABILITY | 246. FLEXIBLE | 298. INNOVATE |
| 195. ENABLE | 247. FOCUS | 299. INNOVATION |
| 196. ENDORSEMENT | 248. FORESEE | 300. INNOVATIVE |
| 197. ENERGY | 249. FORESEEABLE | 301. INSTALL |
| 198. ENHANCE | 250. FORESEEN | 302. INSTALMENTS |
| 199. ENHANCING | 251. FORMER | 303. INSTITUTE |

304. INSTRUCT	355. MINIMUM	407. PATHWAYS
305. INSTRUMENT	356. MINISTRY	408. PEDAGOGIC
306. INTEGRAL	357. MINOR	409. PEDAGOGICAL
307. INTEGRATE	358. MISREPRESENTATION	410. PEDAGOGIES
308. INTEGRATED	359. MOBILE	411. PEER
309. INTEGRATION	360. MOBILITIES	412. PENALTY
310. INTENSE	361. MOBILITY	413. PERCENT
311. INTER	362. MODALITIES	414. PERIOD
312. INTERACT	363. MODIFY	415. PERSPECTIVE
313. INTERCULTURAL	364. MODULES	416. PHASE
314. INTERIM	365. MONITOR	417. PHYSICAL
315. INTERMEDIARY	366. MOTIVE	418. PILOT
316. INTERNAL	367. MULTI	419. PLACEMENT
317. INTERNATIONALISATI ON	368. MULTIDISCIPLINARY	420. PLACEMENTS
318. INTERNSHIPS	369. MULTILATERAL	421. PLATFORM
319. INVEST	370. MULTIPLE	422. POLICY
320. INVOLVE	371. MUTUAL	423. PORTAL
321. ISSUE	372. NARRATIVE	424. POSITIVE
322. JOURNAL	373. NATIONALS	425. POSSESS
323. JOURNALISM	374. NEIGHBOURING	426. POSTMARK
324. JUSTIFY	375. NETWORK	427. POTENTIAL
325. KOREAN	376. NETWORKING	428. POVERTY
326. LABORATORY	377. NEVERTHELESS	429. PRECEDE
327. LABOUR	378. NORMAL	430. PRECISE
328. LAUNCHED	379. NOTIFICATION	431. PREFINANCING
329. LECTURE	380. OBJECTIVE	432. PREJUDICE
330. LEGAL	381. OBJECTIVES	433. PREMISE
331. LEGISLATE	382. OBLIGATION	434. PREPARATORY
332. LEGISLATION	383. OBTAIN	435. PRESIDENT
333. LEXURISERV	384. OCCUPATION	436. PREVIOUS
334. LIABLE	385. OCCUPY	437. PRIMARILY
335. LIFECYCLE	386. ONGOING	438. PRIMARY
336. LITERACY	387. ONLINE	439. PRINCIPLE
337. LOCATE	388. OPTIMAL	440. PRIOR
338. LOGIC	389. OPTION	441. PRIORITY
339. MAINSTREAMING	390. ORDINATING	442. PROCEED
340. MAINTAIN	391. ORGANISER	443. PROCESS
341. MAJOR	392. ORIENT	444. PROCUREMENT
342. MANDATE	393. ORIGIN	445. PROFESSION
343. MATRIX	394. OUTCOME	446. PROFESSIONAL
344. MAXIMISE	395. OUTGOING	447. PROFESSOR
345. MECHANISM	396. OUTLINED	448. PROFILE
346. MECHANISMS	397. OUTPUT	449. PROJECT
347. MEDIA	398. OVERALL	450. PROMOTE
348. MEDIUM	399. OVERSEAS	451. PROOF
349. MEMORANDA	400. OVERVIEW	452. PROSPECT
350. MEMORANDUM	401. PARLIAMENT	453. PROVISION
351. METHOD	402. PARTICIPANT	454. PUBLICATION
352. MIGRANT	403. PARTICIPATE	455. PUBLICATIONS
353. MIGRANTS	404. PARTICIPATING	456. PUBLISH
354. MIGRATE	405. PARTICIPATION	457. PURCHASE
	406. PARTNER	458. QUALITATIVE

- | | | |
|---------------------|---------------------|-----------------------|
| 459. QUANTITATIVE | 508. SELECT | 557. TERRITORY |
| 460. QUANTITY | 509. SEMESTER | 558. TERTIARY |
| 461. RANGE | 510. SEMINAR | 559. THEMATIC |
| 462. RATIO | 511. SENIOR | 560. THEME |
| 463. RECEIPT | 512. SIGNATURE | 561. THEREBY |
| 464. RECOVER | 513. SIGNIFICANT | 562. TIMETABLE |
| 465. RECRUITMENT | 514. SIMILAR | 563. TOPIC |
| 466. RECTOR | 515. SOURCE | 564. TRADITION |
| 467. REFERENCE | 516. SPECIALISED | 565. TRAINEES |
| 468. REFORM | 517. SPECIALISING | 566. TRANSATLANTIC |
| 469. REFORMS | 518. SPECIFIC | 567. TRANSFER |
| 470. REGIO | 519. SPECIFIED | 568. TRANSIT |
| 471. REGION | 520. SPECIFY | 569. TRANSLATE |
| 472. REGISTER | 521. STABLE | 570. TRANSNATIONAL |
| 473. REGULATE | 522. STAKEHOLDERS | 571. TRANSPARENCY |
| 474. REIMBURSED | 523. STATISTIC | 572. TRANSPARENT |
| 475. REIMBURSEMENT | 524. STATUS | 573. TRANSPORT |
| 476. REINFORCE | 525. STATUTORY | 574. TRANSVERSAL |
| 477. REINFORCING | 526. STIMULATE | 575. TRIANGLE |
| 478. REJECT | 527. STRAND | 576. TUITION |
| 479. RELEVANT | 528. STRATEGY | 577. UNDERGO |
| 480. REMUNERATION | 529. STRUCTURE | 578. UNDERGRADUATE |
| 481. REQUIRE | 530. SUBCONTRACTING | 579. UNDERREPRESENTED |
| 482. RESEARCH | 531. SUBMISSION | 580. UNDERTAKE |
| 483. RESIDE | 532. SUBMIT | 581. UNDERTAKEN |
| 484. RESOLUTION | 533. SUBMITTED | 582. UNDERTAKING |
| 485. RESOLVE | 534. SUBPROGRAMME | 583. UNILATERAL |
| 486. RESOURCE | 535. SUBSEQUENT | 584. UPDATED |
| 487. RESPECTIVE | 536. SUBSISTENCE | 585. VALID |
| 488. RESPOND | 537. SUCCESSOR | 586. VALORISATION |
| 489. RETAIN | 538. SUFFICIENT | 587. VARY |
| 490. REVENUE | 539. SUMMARY | 588. VEHICLE |
| 491. REVIEW | 540. SUPERVISE | 589. VERIFICATION |
| 492. ROLE | 541. SUPERVISING | 590. VERIFY |
| 493. SATISFACTORILY | 542. SUPERVISION | 591. VERSION |
| 494. SCAN | 543. SUPPLEMENT | 592. VETERINARY |
| 495. SCANNED | 544. SURPLUS | 593. VIRTUAL |
| 496. SCHEDULE | 545. SUSTAIN | 594. VISIBLE |
| 497. SCHEME | 546. SUSTAINABILITY | 595. VITAL |
| 498. SCHOLAR | 547. SUSTAINABLE | 596. VOCATIONAL |
| 499. SCHOLARSHIP | 548. SYNERGIES | 597. VOLUME |
| 500. SCOPE | 549. TACKLE | 598. VOLUNTARY |
| 501. SCORE | 550. TARGET | 599. VULNERABLE |
| 502. SCORING | 551. TASK | 600. WEBSITE |
| 503. SECTION | 552. TECHNICAL | 601. WORKPACKAGE |
| 504. SECTOR | 553. TECHNOLOGIES | 602. WORKPLACE |
| 505. SECTORAL | 554. TECHNOLOGY | 603. WORKPLAN |
| 506. SECURE | 555. TENDER | 604. WORKSHOP |
| 507. SEEK | 556. TERMINATE | |