Please reference this paper as:

Jablonkai, R. (2020). Leveraging professional wordlists for productive vocabulary knowledge. *ESP Today, 8*(1), 165-181.

Leveraging professional wordlists for productive vocabulary knowledge

Abstract

Productive knowledge of subject-specific vocabulary is essential for successful professional communication. This article puts forward the case for an innovative approach to course and materials design in English for Professional Purposes (EPP) that highlights the importance of careful analysis of the vocabulary of specific professional discourse. It argues that EPP courses would benefit from being informed by corpus-based analysis of vocabulary and collocational choices in texts used in professional contexts. The argument is supported by the results of the corpus-based analysis of the discourse in the professional context of the European Union institutions. The analysis was carried out using the 1-million-word English EU Discourse Corpus (EEUD Corpus), which was created based on a target needs analysis. The present study contributes to knowledge in the field by establishing the first comprehensive EU word and collocation list, which comprises 405 word families and is complemented by collocational patterns specific to English EU discourse. The results underpin the article's central argument that collocational information should be used to enrich professional wordlists as they reveal subject-specific patterns that are fundamental for productive vocabulary knowledge in efficient professional communication. The pedagogic applications of the word and collocation lists are also demonstrated.

Keywords: ESP wordlist, ESP collocation, English for Professional Purposes, English EU discourse, corpus-based analysis,

1. Introduction

The last two decades have seen a surge in corpus-based research into disciplinary vocabulary resulting in wordlists for English for Academic Purposes (e.g. Coxhead, 2000; Dang, 2018; Dang, Coxhead, & Webb, 2017; Gardner & Davies, 2014). Little attention has, however, been devoted to research into vocabulary and wordlists for English for Professional Purposes (EPP), the branch of ESP that "caters for the actual needs of (future) professionals at work" (Ypsilandis & Kantaridou, 2007, p.69). Given that English has become the *lingua franca* in many professional contexts, most notably in international organisations, science and business (Galloway & Rose, 2015), EPP wordlists are crucial especially in the following two educational contexts: one, for in-service English courses of companies and international organisations to improve the English skills of their professionals (Biel, Biernacka, Jopek-Bosiacka, 2018; Nelson, 2006) and two, for ESP courses at universities mainly in non-English speaking contexts to prepare students for their future careers and professions rather than their studies (Ruiz-Garrido, Palmer, Fortanet-Gómez, & Fortanet, 2010; Tangpijaikul, 2014).

Wordlists have been criticised for providing learners with receptive knowledge of vocabulary items, that is, wordlists do not give information on usage patterns and collocations of the individual words (Ackerman & Chen, 2013; Durrant, 2009; Simpson-Vlach & Ellis,

2010; Green & Lambert, 2018; 2019). Therefore, the present study aims to fill this gap by demonstrating how professional wordlists can be compiled and supplemented by collocational information. In addition, the study shows how the identified subject-specific patterns can be applied in EPP instruction directly and indirectly.

The present study extends research into the development and application of wordlists in specific fields by focusing on English language use within the professional context of the European Union institutions. Therefore, a specialised English EU Discourse (EEUD) Corpus was utilised for the purposes of the present study. The design and creation of the EEUD Corpus was based on a target needs analysis carried out among EU experts in relation to their professional contexts. The needs analysis included interviews and a survey with EU professionals to establish the EU documents that they used frequently and felt relevant. The significance of the study is twofold: first, it proposes a way to inform professional wordlist compilation by a target needs analysis; second, it shows how professional wordlists can be supplemented by collocational information and can be used to develop teaching materials.

The article begins with a critical review of previous studies on wordlists, subject-specific collocations and English EU discourse. It then presents the methodological approach, including the procedures of corpus creation, and criteria for word and collocate selection. Following that, the findings are discussed. These highlight the relevance of the specificity of wordlists for EPP learning. Then some recommendations regarding the pedagogic applications of the EUWL are offered. The conclusion argues that EPP wordlists should be established based on careful analysis of the professional discourse, ideally based on a target needs analysis and should include collocational information to provide productive knowledge of technical and highly frequent vocabulary.

2. Wordlists in ESP

A neglected area within ESP vocabulary studies is the analysis of the vocabulary of different professions. The majority of previous research into subject-specific vocabulary have investigated academic disciplines and compiled wordlists, for example, for hard and soft sciences, such as Physics and Education, or for EAP courses at universities (e.g. Coxhead, 2000; Dang, 2018; Dang, Coxhead, & Webb, 2017; Gardner & Davies, 2014). More recently this kind of vocabulary research has been extended to the compilation of wordlists for trade education (Coxhead & Demecheleer, 2018) and for academic literacy in secondary education (Green & Lambert, 2018; 2019). Most of these wordlists were compiled with the aim to support the learners' studies in different educational contexts. However, in order to provide EPP learners with the tools for effective and efficient communication in English in their careers, similar efforts must be made to teach them the technical and highly frequent vocabulary of their professions as used in their respective professional contexts (Ypsilandis & Kantaridou, 2007).

With this aim in mind, it is essential that the corpora created to compile wordlists include texts that represent the discourse that is relevant in the current or future professional contexts learners are or will be working in (Nation, 2016). In this respect, there are two main limitations of previous studies regarding the corpora they utilised: one, the corpora included textbooks and research articles of their respective disciplines that represent the academic rather than the professional variety of English discourse (e.g. Bi, 2020; Dang, 2018; Lei & Liu, 2016; Yang, 2015); and two, the selection of texts was rarely based on careful target needs analysis and the systematic collection and analysis of texts and their use in the relevant professional contexts (Nelson, 2006). Furthermore, the few studies that have investigated professional discourses (Biel et al., 2018; Freund, 2014; Tangpijaikul, 2014; Trebits, 2008; 2009a; 2009b) analysed corpora that were compiled based on either the researcher's intuition (e.g.

Tangpijaikul, 2014; Trebits, 2008) or on the advice of a limited number of experts in the professional field (Biel et al., 2018). In order to adequately inform the course and materials design process, a more systematic target situation analysis is needed. The present study contributes to the field by demonstrating how a target needs analysis can inform the corpus compilation.

Studies have demonstrated that "a pedagogical focus on productive vocabulary is at least as important as one on receptive vocabulary" (Durrant, 2016: 50). However, a major limitation of wordlists is that they do not provide phraseological and lexico-grammatical information, as they only include single-word units (Simpson-Vlach & Ellis, 2010; Green & Lambert, 2018; 2019). This makes them more suitable to teach receptive (reading and listening) rather than productive vocabulary knowledge (writing and speaking) (Nation, 2016). Productive knowledge of a word requires knowledge and mastery of the following aspects of use: (1) what patterns the word is used in; (2) what words the word is used together with; and (3) what registers, subject fields, etc. can the word be used in (Nation & Hunston, 2018). The present study argues that an effective way to present this knowledge to language learners is to complement single-word unit wordlists with collocational information. Furthermore, in order to provide productive knowledge of technical and semi-technical words, learners need to be shown how these collocational choices are unique in their specific professional fields (Nelson, 2006; Walker, 2011). Previous studies have demonstrated that the collocational frameworks of technical words are subject-specific and a good command of the collocational patterns that are typical of the language use of the professional field is necessary in order to communicate effectively in a profession (Bartsch, 2004; Nelson, 2006).

This study contributes to the literature of ESP vocabulary studies by developing a wordlist for an under-researched professional field supplemented by collocational information. It is argued that enriching wordlists with collocational information and providing EPP learners with subject-specific collocational patterns of technical and highly frequent vocabulary items is essential for mastering productive knowledge of these words in order to facilitate effective and efficient professional communication.

3. **Identifying pedagogically relevant subject-specific vocabulary for the profession**

Subject-specific or technical vocabulary is defined as the words that are closely associated with a subject field (Nation & Hunston, 2018: 303) and have a specific meaning in the field (Ha & Hyland, 2017). Although they can come from all three frequency levels of vocabulary (high, medium, and low), high-frequency subject-specific vocabulary is typically considered pedagogically relevant for wordlists (Nation, 2016). Previous studies into subject-specific vocabulary and term recognition suggest that quantitative and qualitative selection criteria are necessary to reliably identify these words (Nation, 2016; Marín, 2014; Kwary, 2011). Marín, for example, compared five automatic term recognition methods and found that none of the methods in her analysis identified more than 73.45% of the pre-defined list of legal terms in her specialized legal corpus. The keyword analysis method (Scott, 2008), which compares frequencies of words in a target and a reference corpus and determines the words that are unusually frequent in the target corpus, recognised 62% of the legal terms. Marín concludes that some kind of qualitative method, for example, consultation with subject specialists is necessary in order to disambiguate words that have several meanings. Such corpus-comparison approaches have two further weaknesses. First, they typically compare frequencies of word forms which is not very meaningful for pedagogical purposes. Second, they do not take into consideration the range of words, that is, how frequently they are used in individual texts or sub-corpora within the specialised corpus. Nation (2016), however, suggested that the most important quantitative criterion for including words in a pedagogical wordlist was their range as

it shows how widely the word is used. This criterion is also an especially important consideration for the present analysis of English EU discourse as it aims to identify subject-specific vocabulary used in texts representing the different EU fields of activity.

Additional quantitative selection criteria for identifying subject-specific words proposed in the literature were specialised occurrence and frequency (Coxhead, 2000; Nation, 2016). Specialised occurrence is typically ensured by excluding general words, such as the words of the General Service List (GSL, West, 1953) or the most frequent 2000 words in the British National Corpus (BNC 2000, Nation, 2004), from among the frequently occurring word families in a specialised corpus (e.g. Coxhead, 2000; Hsu, 2013; Liu & Han, 2015; Yang, 2015). In addition, the selection of word families is often guided by a minimum cumulative frequency of occurrence of a word family. The level of minimum cumulative frequency is usually set using Coxhead's (2000) 100 occurrences in a 3.5 million-word corpus as a benchmark adjusting the frequency count to the size of their respective corpora assuming a linear relationship between corpus size and the number of word types in a corpus (e.g. Yang, 2015). Overall, previous studies applied varying quantitative, primarily frequency-based selection criteria to develop pedagogical subject-specific wordlists and many argued that a combination of quantitative and qualitative selection methods are needed.

4. Analysis of vocabulary in written English EU discourse

The present study focuses on the written English professional discourse in EU institutions. In the last couple of decades, English has gained prominence in EU institutions as *lingua franca*, therefore, it is crucial that EU professionals whose first language is not English have excellent English skills for professional communication in decision and policy making (Fischer, 2010; Galloway & Rose, 2015; Truchot, 2002). Despite the fact that the UK has left the EU, it is highly likely that English will continue to play an important role when negotiating new policies and drafting EU documents and will, thus, in all likelihood still remain a powerful language as a *lingua franca* in many EU contexts (Ginsburgha, Moreno-Ternero, and Weber, 2017; Modiano, 2017).

A unique feature of English EU discourse from a vocabulary point of view is that the European Union is active in a very wide range of topics and EU institutions produce documents in areas such as agriculture, customs, trade, budget, education, and research. The present study aims to capture the EU-specific vocabulary that represents concepts, procedures, and communication that are relevant when working in the institutions of the European Union in general and not only in one specific topic area. Therefore, texts produced by EU institutions relating to all these topics were included in the corpus used by this study.

The handful of studies that have so far investigated official English EU texts explicitly for pedagogic purposes focused on very specific genres and registers, for example, grant calls (Freund, 2014) or one specific topic area, such as EU Competition Law (Biel et al., 2018). Trebits (2008; 2009a; 2009b) analysed a very small corpus of 200,000 words of English EU documents (information booklets, annual general reports and sample EU recruitment test) that was compiled intuitively, revealing that 46.5% of the word types are not among the BNC 3000 (the first 3,000 most frequent words of the British National Corpus, Nation, 2004). This suggests that a substantial number of vocabulary items in English EU texts are not part of the vocabulary of an intermediate level language learner (B1- B2 levels according to the Common European Framework of Reference, Alderson, 2002) and highlights the importance of comprehensive analysis of the vocabulary of official English EU texts for EPP pedagogic purposes (Trebits, 2008; 2009a; 2009b). Therefore, as part of a larger project that investigated the variety of English used in official EU documents, this study was undertaken to analyse the vocabulary in English EU documents to establish a wordlist of EU-specific vocabulary and to identify EU-specific collocational patterns that can inform course and materials design to

facilitate productive subject-specific vocabulary knowledge. The goal was to cater for learners' target situation needs and analyse the vocabulary of texts they will use in their professional contexts. Therefore, the following research questions were formulated to guide this analysis:

(1) Which vocabulary items occur frequently in the written English EU discourse and can be considered as pedagogically relevant subject-specific words?
(2) To what extent are collocational patterns in English EU discourse subject-specific?
(3) How can the findings in the present study inform EPP instruction?

5. Methods

5.1. The English EU Discourse Corpus

The corpus created for the study contained 1,174,753 running words from 241 written texts representing 40 different EU genres, such as treaties, regulations, press releases, presidency conclusions, calls for proposals (see Table 1). During the corpus design and creation, great care was taken to develop a reliable corpus that is representative and balanced (Biber, 1993). Therefore, to ensure that the corpus comprises texts that represent the discourse that is characteristic of learners' present or future professional contexts the corpus building process was based on a needs analysis survey among EU professionals who worked in one of the EU institutions or EU-related governmental bodies to identify the relevant EU genres and EU documents for sampling. In Phase 1, of this target needs analysis, interviews with 10 EU professionals were conducted. EU professionals included EU experts, translators and interns who worked at the EU Commission and at the Hungarian EuroDirect, the EU information service of EU issues to the general public. In Phase 2, an online questionnaire was administered among EU professionals who worked in EU institutions and EU-related bodies in the Hungarian government. The 99 respondents identified specific texts and genres they used in their work, indicated the relevance of specific texts and genres in their jobs and how frequently and for what purposes they used them in their daily work.

The survey results regarding frequency and relevance of use determined the proportion of different genres in the corpus. This served as a sampling frame as proposed by Biber (1993) for more representative corpus building. The detailed contents of the corpus listing the different genres that were included can be found in Table 1. Another important factor in corpus design is balance (Biber, 1993). As the focus in the present study was to identify vocabulary associated with the EU in general, and not with one specific EU field of activity, efforts were made to balance the corpus for the different fields of EU activities, for example, economy, agriculture, security policy, education, and single market (Jablonkai, 2010a). There were altogether 34 sub-corpora created according to the EU fields of activity defined according to the list of EU policies available on the official website of the EU (EU website, 2018). Only texts published by one of the EU institutions, for example, the Commission, the Parliament, and the Council were included in the corpus. The sample EU texts were kept at their original length, but the reference sections where different pieces of EU legislation were listed were deleted.

Text categories	Genres	Length (number of words)	Number of texts	% of corpus
EU legal texts	Treaties, International agreements, Regulations, Directives, Decisions, Recommendations, Opinions, Common positions CFSP, Judgements of the Court of Justice	521,554	81	44.5%
Legislative preparatory documents	Commission legislative proposals, Council's common positions,	217,894	42	18.5%

	Legislative resolutions of the European Parliament, Commission communications, Green papers, White papers, ECOSOC Opinions, EP Positions, EP Draft Reports, EP initiatives			
Documents related to EU funds	Calls for proposals, Application forms, Project contracts, Ex_ante_guides, Grant agreements, Guide for applicants, Project fiches	118,144	24	10%
Other documents issued by EU institutions	Commission Working Documents, Rules of procedures, Press releases, Resolutions, Declarations, Presidency conclusions, Community guidelines, Common strategies, Commission Notices, Presidency Notes, Council minutes and addenda to minutes, Press conferences, Operation manuals, Reports	317,161	94	27%
Total		**1,174,753**	**241**	**100%**

Table 1. Contents of the written English EU Discourse Corpus (adapted from Jablonkai, 2010b: 256)

5.2. Developing the English EU wordlist

The word family (Nation, 2016) was adopted as the unit of analysis for the purposes of compiling the EU wordlist for three reasons. First, the target learners of EU English courses often have an intermediate (B2) level of English proficiency and have some morphological and word building skills to benefit from the transparency of word families (Nagy et al., 1989). Second, ESP wordlists arranged by word families can also be used to raise learners' morphological awareness as they include subject-specific affixes and suffixes (Bauer & Nation 1993; Nation, 2016). Finally, this will make the EUWL comparable to earlier analyses of ESP and general vocabulary as many of these wordlists are organised around word families (Freund, 2014; Nation, 2016; Tongpoon-Patanasorn, 2018). At the same time, a lemma list version of the EUWL was also created for the purposes of the collocational analysis.

To develop the EUWL, the corpus analysis programmes Range (Heatley et al., 2002) and Wordsmith Tools (Scott, 2008) were used. Wordsmith Tools was used to generate the initial frequency list and to run a keyword analysis with the general BNC World corpus as the reference corpus using the log likelihood statistic and a frequency threshold of 3. Next, the keyword list was organised into word families by the function of Wordsmith that merges certain entries according to a pre-prepared list. The Range programme was used to measure the range of word families by counting the frequency of word types in the individual sub-corpora and record the frequency of occurrence of individual word types in total and in each sub-corpus.

Three quantitative selection criteria were adopted in this study: specialised occurrence, range and cumulative frequency (Nation, 2016). First, specialised occurrence was ensured by eliminating the most frequent 2000 word families as represented by the BNC/COCA list (Nation, 2017) from among the word families developed from the keyword list. The BNC/COCA list was used as it is the latest general wordlist and it is organised by word families (Nation, 2016). Second, only word families used in a wide range of EU fields of activity were

selected to ensure that the wordlist is EU-specific and balanced for the different EU fields of activity. Word families had to occur in 16 or more of the 34 EU-related fields of activity. Third, this study started out from the cumulative frequency criterion set by Coxhead (2000) at 100 in her 3.5-million-word corpus as a benchmark for many wordlists (Nation, 2016). Taking the non-linear relationship between corpus size and the number of word types in a corpus into consideration, however, the present study applied Biber's (2006) simple formula to adjust the number of word types in corpora of different sizes. According to Biber's findings half a corpus represents around 70% of the word types in the larger corpus. His formula says that the ratio of the number of word types in two corpora (e.g. 0.7) is the square root of the ratio of the number of total running words in the two corpora (e.g. 0.5). The same formula should be applied to setting the threshold for word selection. The corpus used by Coxhead was three times bigger than the corpus used in this study. Therefore, the adjusted cumulative frequency threshold for inclusion into the EUWL was set at 57, as the square root of one-third is 0.57.

In order to ensure the quality and relevance of the EUWL, the quantitative criteria were combined with qualitative criteria in the selection process. Therefore, the final step of establishing the EUWL involved two subject specialists to clarify ambiguous cases (Coxhead & Demecheleer, 2018; Tongpoon-Patanasorn, 2018). One of them was an EU expert and the other one was an ESP teacher with experience to teach English in the institutions of the European Union. Vocabulary items were included in the final EUWL if both experts found that all of the following four requirements were met: a) the meaning of the word is related to the field and should be taught; b) the meaning of the word is related to the field and EU professionals should know this word; c) the word has a subject-specific meaning and should be taught; and d) the word has a subject-specific meaning and EU professionals should know this word.

To evaluate to what extent the list is subject-specific and to establish its added usefulness for pedagogical purposes, the text coverage, – that is, the instances of words in a corpus that are covered by the elements of a wordlist (Nation & Kyongho, 1995) – of the final EUWL was tested in several registers and genres from different sources as recommended by Nation (2016). This validation of the final EUWL was carried out with the help of the Range programme (Heatley et al., 2002).

5.3. Collocation analysis

In order to go beyond mere lists of collocates and to present more detailed collocational patterns of the vocabulary items in the EUWL, the present study applied quantitative and qualitative methods to analyse the collocational frameworks of the words of the EUWL. The concept of collocation was introduced by Firth (1968) and it was elaborated by Sinclair, who defined collocation as "the occurrence of two or more words within a short space of each other in a text" (Sinclair, 1991 p. 170). The quantitative analysis was conducted with the help of Wordsmith Tools (Scott, 2008) and the following selection criteria were applied:

1. Statistical measures: Mutual Information (MI) score of 4 or higher. MI measures the strength of association between pairs of words. The most commonly used threshold is 3 to indicate a meaningful relationship (Lei & Liu, 2018). However, recent studies that tested frequency-based methods to identify collocations suggested that MI scores higher than 3 result in psychologically real collocates (Durrant & Doherty, 2010). Therefore, the present study applied the cut-off point of MI score 4 for including collocates.
2. Minimum frequency: more than 5 co-occurrences within a 4-word span. MI scores tend to give undue weight to low frequency words and eliminate words that frequently co-occur with many words (e.g. the) (Lei & Liu, 2018). Therefore, a minimum number of co-occurrence threshold was set at 5 within 4 words to the left or right of the EUWL word (Walker, 2011).

3. Range: collocations had to occur in at least 10% of all texts in the EEUD corpus. This measure was applied to ensure that the collocation was used across several texts and does not represent the idiosyncratic language use of a single text.

Previous studies found that word forms and lemmas display different collocational patterns in corpora (Tognini-Bonelli, 2001; Hoey, 2005). Although investigating collocational patterns of individual word forms is interesting for linguistic purposes, as the aims of the present study were primarily pedagogical, examining lemmas was found to be more appropriate as this provides an adequate level of detail to language learners (Nation, 2016). Each inflectional form of a lemma was included in the search for collocations.

The next step in collocational analysis was to compare collocational patterns in the specialised corpus to the ones in a general corpus, the BNC written, with the help of Sketch Engine. In addition to analysing word sketches of selected frequent vocabulary items, the semantic preferences that emerged from the collocates were also identified by qualitatively analysing their concordance lines (Nelson, 2006; Stubbs, 2001). Stubbs (2001) defined semantic preference as "the relation, not between individual words, but between a lemma or word-form and a set of semantically related words" (p. 65). The semantic preferences of words can inform us about the different shades of meaning they can express as well as the context of the language use (Nelson, 2006). The patterns in the EEUD Corpus were compared to patterns identified in the written section of the BNC. First, the collocates in the same grammatical relations were grouped into relevant semantic sets and summarised in a table format as illustrated in Table 5 with the data of the lemma CRITERION. Next, the identified preferential semantic sets were compared across the general and the specialised corpora. For the purposes of this analysis 12 lemmas from the EUWL were selected based on their pedagogical value. The list included six nouns: *policy, commission, criterion, regulation, initiative, objective*; two adjectives: *European; eligible* and four verbs: *notify, function, ensure, implement*.

6. Results and Discussion

6.1. Elements of the EU wordlist

The final EUWL contains 405 word families that are made up of 1,898 word types and 611 lemmas. Table 2 gives an example of the word families with its members in the EUWL. The word families among the most frequent ones include *EUROPE, COMMISSION, REGULATION* and *IMPLEMENT.* Examples of the least frequent word families are *CAMPAIGN, VULNERABLE, WORLDWIDE, HIGHLIGHT* and *ALIGN*. The EUWL includes word families in connection with funding such as *BENEFICIARY* and *RESOURCE*, the main EU institutions such as *COMMISSION, PARLIAMENT* and *PRESIDENCY*, and legal words such as *REGULATE* and *TREATY*. In addition, the wordlist contains abbreviations, for example, *DG, EC, OJ, SME* and geographical names, such as all member states and names of two cities: *BRUSSELS* and *LISBON*, and a few function words, such as *PRIOR, BEHALF* and *VIA*. The headwords of the word families in the final EUWL are given in the Appendix.

N	Headword	Cumulative frequency	%	Members of the word family
1	EUROPEAN	7401	0.69%	europe[600] europe's[90] cross-europe[1] e-europe[11] european[6621] european-based[1] european-wide[1] europeans[29] intra-european[3] non-european[20] trans-european[23] transeuropean [1]

Table 2. Example of an EU word family

6.2. Subject-specificity of the EUWL

The EUWL was tested for its specificity for EU discourse and relevance for English for EU pedagogic purposes by measuring its coverage of texts representing different registers and genres. As shown in Table 3, the EUWL accounts for 14.06% of the tokens in the EEUD Corpus. The EUWL reached a high coverage – 13.10% – of another corpus of EU texts, which was compiled according to different selection criteria than the EEUD Corpus (Trebits, 2009a). Thus, the high coverage reinforces the validity of the EUWL as a wordlist useful for understanding English EU texts in general.

Texts	Tokens	Text coverage
EEUD Corpus	1,076,460	14.06%
EU English Corpus	197,620	13.10%
20th century literary texts	105,578	0.88%
News texts	117,164	4.76%

Table 3. Text coverage of EUWL in different genres and registers

In order to establish whether the EUWL is a truly EU-specific wordlist, it was also tested on literary texts and news texts. As can be seen in Table 3, the elements of the EUWL accounted for 0.88% in literary texts. Not surprisingly, this register seems to be very different from the EU discourse regarding its vocabulary. News texts with slightly less than 5% coverage also seem to use a markedly different vocabulary from EU texts. This highlights that although it is often common practice in EPP courses to use news texts, most probably because of their relatively easy access, news texts might exhibit a very different language variety than the texts used in professional contexts. Therefore, the results of the present study question the appropriateness of using news texts in EPP courses of EU English. They support the argument that EPP courses should be informed by specific professional word and collocation lists that are compiled based on the analysis of the professional discourse represented by texts used by professionals at work and suggest that EPP courses should use such authentic professional texts.

6.3. Results of the collocation analysis

An extract from the collocation list of the lemmas of the EUWL is presented in Table 4. As a novel approach, the present study extended the investigation to gain insights into the subject-specific nature of collocates by comparing collocational frameworks between the specialised EEUD corpus and the general BNC Written corpus.

European (adj. 6742)*	commission (n. 5070)	implement (v. 1005)	criterion (n 370)
parliament (1150/12.78)**	proposal (258/6.02)	measure (152/11.37)	eligibility (21/11.42)
union (987/12.6)	inform (86/6.28)	programme (37/10.08)	award (18/11.22)
council 460/11.48)	communication (156/5.89)	rule (33/10.04)	selection (19/11.16)
community (220/10.62)	communities (53/5.09		set (34/10.46)
commission (165/10.62)	report (117/4.88)		follow (23/9.34)
bank (151/10.1)			

Table 4. Example EU-specific collocations ordered by MI score
*Total frequency
** (Frequency of co-occurrence/MI)

6.4. Subject-specificity of the collocational patterns

The findings reveal that the collocational patterns of the investigated lemmas are subject-specific to some extent. The comparison of the collocates in the general and the EEUD corpus shows that the collocations in the EEUD Corpus suggest a higher degree of fixedness (Gledhill, 2000; Nelson, 2006), that is, the proportion of collocates of individual lemmas covered by semantic preferences is higher in the specialised corpus. The greater number of semantic sets identified among the collocates of the lemmas in the BNC Written also supports the concept of fixedness in collocational patterns in a specialised corpus. In the case of the selected lemmas the number of semantic sets ranges from 5 to13 in the EEUD Corpus and 7 to 23 in the BNC Written.

The findings of the present study confirm that the words in a specialised corpus are associated with subject-specific semantic preferences and also with semantic sets that are the same in both in the specialised and the general corpus of English (Nelson, 2006). The comparison of the number of identical semantic sets that lemmas are associated with in the two corpora, shows that the analysed lemmas have in general 2-6 identical semantic sets. The highest number of identical preferential semantic sets were identified in the case of CRITERION and the lowest one in the case of the lemma EUROPEAN. Table 5 presents the comparison of the collocates of CRITERION in the two corpora. The collocates are grouped according to the respective semantic sets within the identified grammatical relation categories. For example, the first semantic set in the grammatical relation: 'object of' comprises the collocates that mean 'to meet a criterion', which in the EEUD corpus are: *fulfil, fulfill, meet, satisfy.*

CRITERION noun	
BNC Written	EEUD
grammatical relation: object of	
Semantic set 1 meet collocates: *satisfy, fulfil, meet, match, fit* Semantic set 2 set collocates: *formulate, adopt, outline, define, establish* Semantic set 3 respect collocates: - Semantic set 4 list collocates: *list, specify* Semantic set 5 apply collocates: *apply, use, employ* Semantic set 6 evaluate collocates: *assess, judge, review*	Semantic set 1 meet collocates: *fulfil, fulfill, meet, satisfy* Semantic set 2 set collocates: *set, agree, establish, lay* Semantic set 3 respect collocates: *follow, respect* Semantic set 4 list collocates: *list, specify, give* Semantic set 5 apply collocates: *apply* Semantic set 6 evaluate collocates: -
Other collocates: *invoke, exemplify, propose, interpret, identify, derive, alter*	Other collocates: *need, see, base, propose*
Number of preferential semantic sets	
5	5
Number of identical preferential semantic sets: 4	
grammatical relation: pp for	
Semantic set 7 participation collocates: *eligibility, inclusion, exclusion* Semantic set 8 evaluation collocates: *selection, evaluation, assessment, diagnosis* Semantic set 9 membership collocates: *admission, acceptance, membership,*	Semantic set 7 participation collocates: - Semantic set 8 evaluation collocates: *selection* Semantic set 9 membership collocates: *membership*

entry, access Semantic set 10 distribution of funds collocates: -	 Semantic set 10 distribution of funds collocates: *allocation*
Other collocates: *imposition, promotion, recognition, transfer, success, use, service*	Other collocates: *Websites*
Number of preferential semantic sets	
3	3
Number of identical preferential semantic sets: 2	
Total number of preferential semantic sets:	
8	8
Total number of identical preferential semantic sets: 6	

Table 5. Comparison of semantic preferences of the lemma CRITERION

Overall, the results of the present study confirm the EUWL as a list of subject-specific words that can be considered as pedagogically relevant. Furthermore, the findings of the collocation analysis indicate that the selected lemmas demonstrate subject-specific collocational patterns that are, in some cases, markedly different from patterns in the general corpus. It should be noted, however, that a much wider collocational analysis would be necessary to make definite claims regarding the nature of these differences. Nevertheless, there seems to be enough evidence to support the argument for the necessity of professional wordlists for EPP purposes and the importance of supplementing professional wordlists with subject-specific collocational information.

7. Pedagogical implications

The findings of the present study can be used for pedagogy directly and indirectly. Indirectly, the EUWL and collocation list can serve as a firm basis for course and materials design. A strong argument for the application of the EUWL in EPP teaching is the high coverage of English EU texts it provides. As it is shown in Table 6, the first 2000 word families of the BNC/COCA list and the families of the EUWL together account for 92.13% of the EEUD corpus, which is higher than the coverage of the general BNC/COCA3000 wordlist. As a result, the EUWL with its subject-specific elements helps learners reach closer to the level of 98% coverage which is considered necessary for understanding a text without a dictionary (Hirsh & Nation, 1992; Nation & Waring, 1997). The evaluation of the EUWL also demonstrated that it is subject-specific and comprises word families that are used in a wide range of EU texts. It can also provide guidelines for the sequencing of the teaching of vocabulary items, as teaching can follow the frequency order of the word families in the list. With the help of the EUWL, the EU-specific elements can easily be selected and can be used as the basis for traditional vocabulary teaching exercises as well as for data-driven learning activities.

Wordlists	Coverage of EEUD Corpus
BNC/COCA 1000	62.75%
BNC/COCA 2000	15.32%
EUWL	14.06%
BNC/COCA1000+2000+EUWL	92.13%
BNC/COCA1000+2000+3000	90.39%

Table 6. Text coverage of general wordlists and the EUWL

Furthermore, given the efficiency of language-focused learning (Nation & Hunston, 2018), the wordlist and collocation list created as part of the present study can be used in

teaching directly. Results of the collocational analysis help learners master productive knowledge of individual vocabulary items (Nation & Hunston, 2018). These can be presented to learners in the following ways: (1) as a list when teaching specific vocabulary items, as presented in Table 4; (2) in the form of pedagogic collocational profiles as shown in Table 7; as a novel element, this profile not only gives language learners guidance on relevant collocates, but it also presents frequent semantic preferences and grammatical relations the particular lemma frequently forms with relevant collocates extending the learners' understanding of the semantic and grammatical patterns of the specific professional discourse; and (3) a comparison of the collocational patterns of the same word can be shown in specialised and general corpora, as can be seen in Table 5 (Nelson, 2006; Walker, 2011). This comparison will heighten the learners' awareness of various features of the language use in professional discourse and will provide them guidance on how to disambiguate slight, but significant differences in meaning and to identify the different uses of a word. Finally, two activity types are presented below to demonstrate how the findings of the present study can be turned into classroom activities for EPP pedagogy.

Activity types

Aim: to raise learners' awareness of collocates of particular vocabulary items
A.1 Instruction: Study the collocational profile of the verb IMPLEMENT in Table 7 and underline the nouns in the table that are likely to be used with it in EU documents.

the accession criteria	Function
Opinion	the acquis
a reform	Measures
the internal market	a directive
Legislation	a summit
a timetable	Policies
a programme	a debate

IMPLEMENT verb	
Construction	**semantic groups**
IMPLEMENT + noun	1. <u>legislation</u> collocates: *measure, rule, regulation, provision, directive, legislation, recommendation, decision, convention* The Commission shall **implement** this Regulation in accordance with the Financial Regulation. 2. <u>plans</u> collocates: *reform, strategy, programme, project, policy, commitment, budget, plan* Many European policies and programmes are **implemented** at regional and local levels. 3. <u>approach</u> collocates: *approach, principle* The forthcoming proposal for a new Directive **implementing** the principle of equal treatment outside employment will be addressed. 4. <u>activity</u> collocates: *action, tool, operation* By way of derogation from paragraph 1 , in-kind contributions , depreciation costs and overheads may be treated as expenditure paid by beneficiaries in **implementing** operations under the following conditions:
IMPLEMENT + adverb	1. <u>positive</u> collocates: *properly, effectively, fully, successfully, actively* The Commission, in its role of guardian of the Treaty, is responsible for ensuring that Community legislation is properly transposed into national law

	and properly **implemented** and enforced by national authorities in the Member States. 2. negative *not* However, Albanian legislation does not yet protect these rights sufficiently and is not fully **implemented**. Other collocates: *systematically, as, directly*

Table 7. Extract from the collocational profile of the verb IMPLEMENT

A.2 Instruction: Study the collocational profile of the verb IMPLEMENT in Table 7 and add five more nouns that are often used together with it in EU documents.

1) 2) 3) 4)
5)

Key to Activity type A.1:

the accession criteria	Function
Opinion	the acquis
a reform	**Measures**
the internal market	**a directive**
Legislation	a summit
a timetable	**Policies**
a programme	a debate

Key to Activity type A.2: a. rule, b. convention, c. plan, d. action (for further examples see Table 7)

To argue for the relevance of EPP wordlists, the present study focused on the professional context of the institutions of the European Union. As the use of English has increased in internal communication within EU institutions, especially, in written communication (Truchot, 2002) in the last couple of decades, the findings of this study can be useful for (1) universities in Europe that offer courses in EU English in their international relations studies and translation programmes (Corvinus, 2018; ELTE, 2018; NKI, 2018; TEMPUS, 2020); (2) published textbooks on EU English (e.g. Trebits & Fischer, 2010); (3) preparing for tests for selecting applicants for positions in EU institutions (which can always be taken in English but not necessarily in other official languages) (EPSO Sample tests, 2019); (4) in-service EU English courses offered to EU professionals (e.g. Pardillos & Ángel, 2016).

8. Conclusion

The article's significance lies in the attempt to leverage the affordances of corpus-informed wordlists for EPP productive vocabulary knowledge for learning and pedagogy. It contributes to our knowledge of ESP vocabulary by investigating a professional language variety where English is used as a *lingua franca*. Furthermore, it adds to the literature by identifying and presenting subject-specific collocational information in order to help EPP learners master productive knowledge of relevant technical and highly frequent words. To confirm the findings of this study a much wider analysis of subject-specific collocational patterns and semantic preferences would be needed that includes the examination of a higher number of lemmas as well as the analysis of other professional fields. The present study has also demonstrated how the creation of corpora for EPP purposes can and should be informed by target needs analysis. This approach is crucial to ensure that pedagogic professional corpora better reflect the language variety in the specific professional context and as such make

wordlists more valid. It should be noted, however, that the findings of the present study primarily refer to written communication within the EU context and, consequently, a limitation of the analysis is its exclusive focus on written English EU discourse. Further research is thus needed into oral communication in English within EU institutions in order to complement findings of the present study by the distinctive characteristics of the spoken professional discourse.

References

Ackerman, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235-247.

Alderson, J. C. (Ed.) (2002). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies*. Strasbourg: Council of Europe. Retrieved 28 April, 2016 from http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp

Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: unter Narr.

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

Bi, J. (2020). How large a vocabulary do Chinese computer science undergraduates need to read English-medium specialist textbooks? *English for Specific Purposes, 58*, 77-89.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243-257.

Biber, D. (2006). *University language a corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biel, L. Biernacka, A., & Jopek-Bosiacka, A. (2018). The glossary of EU English Competition collocations and terms. In S. Marino, L. Biel Lucja, M. Bajcic, & V. Sosoni (Eds.). *Language and law* (pp. 249-274). Heidelberg: Springer.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System, 32*, 251-263.

Corvinus (2018). *English for the European Union course description*. Corvinus University of Budapest. Retrieved 15 March, 2018 from http://www.uni-corvinus.hu/index.php?id=26671&tipus=tt&uid=52574702&l=hu

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes, 51*, 84-97.

Dang, T.N.Y. (2018). The nature of vocabulary in academic speech of hard and soft-sciences, *English for Specific Purposes*, *51*, 69-83.

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The Academic Spoken Word List. *Language Learning, 67*(4), 959–997.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157-169.

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes, 43*, 49-61.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory, 6*(2), 125-155.

ELTE (2018). EU English course description, In *BA in English Language Programme handbook*. ELTE University Budapest, p. 24. Retrieved 15 March, 2018 from http://savaria.elte.hu/btk/flli/an/Dokumentumok/tant%C3%A1rgyi%20programok.pdf
EPSO Sample tests (2019). Retrieved 15 May, 2019 from https://epso.europa.eu/how-to-apply/sample-tests_en
EU website (2018). Topics of the European Union, Retrieved 15 March, 2018 from http://europa.eu/pol/index_en.htm
Firth, J. R. (1968). A synopsis of linguistic theory, 1930-55. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth 1952*-59 (pp. 168-205). London: Longmans.
Fischer, M. (2010). Translation(policy) and terminology in the European Union. In M. Thelen, & F. Steurs (Eds.), *Terminology in everyday life*. (pp. 21-34). Amsterdam: John Benjamins.
Freund, Zs. (2014). A corpus analysis of grant guidelines: the Education and Training Programme word list (ETPWL). *Journal of Teaching English for Specific and Academic Purposes, 2*(3), 501-514. Retrieved 30 March, 2015 from http://espeap.junis.ni.ac.rs/index.php/espeap/article/view/123
Galloway, N., & Rose, H. (2015). *Introducing Global Englishes*. New York: Routledge.
Gardner, D., & Davies, M. (2014). A New Academic Vocabulary List. *Applied Linguistics, 35*(3), 305-327.
Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language, 23*, 65-83.
Ginsburgha, V., Moreno-Ternero, J. D., & Weber, S. (2017). Ranking languages in the European Union: Before and after Brexit. *European Economic Review 93*, 139-151.
Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes, 19*(2), 115-135.
Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes, 35*, 105-115.
Green, C., & Lambert, J. (2019). Position vectors, homologous chromosomes and gamma rays: Promoting disciplinary literacy through Secondary Phrase Lists. *English for Specific Purposes, 53*,1-12.
Ha, A. & Hyland, K. (2017). What is technicality? A Technicality Analysis Model for EAP vocabulary. *Journal of English for Academic Purposes, 28*, 35-49.
Heatley, A., Nation, P., & Coxhead, A. (2002). *Range and Frequency programs.* Retrieved 5 May, 2015 from http://www.victoria.ac.nz/lals/about/staff/paul-nation.
Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689-696.
Hoey, M. (2005). *Lexical priming*. New York: Routledge.
Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical wordlist. *Language Teaching Research, 17*(4), 454-484.
Jablonkai (2010a). Towards an English EU Discourse Corpus. In T. Frank, & K. Károly (Eds.). *Gateways to English*. pp. 231-247. Budapest: Eötvös University Press.
Jablonkai (2010b). English *in the context* of European integration: a corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes. Special Issue: ESP in Europe, 29*(4) 253-267.
Kwary, D. (2011). A hybrid method for determining technical vocabulary. *System, 39*(2),175-185.
Lei, L. & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes, 22,* 42-53.

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics, 23*(2)*,* 216-243.
Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes, 39*, 1-11.
Modiano, M. (2017). English in a post-Brexit European Union. *World Englishes, 36*(3), 313–327.
Nagy, W., Anderson, R., Schommer, M., Scott, J.A., & Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, *24*, 262–281.
Nation, P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a second language*. (pp. 3-13). Amsterdam: John Benjamins.
Nation, I. S. P. (2016). *Making and using word lists for language learning and testing.* Amsterdam: John Benjamins.
Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Available from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx
Nation, P. & Hunston, S. (2018). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.
Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy*. Cambridge: Cambridge University Press, 6-19.
Nelson, M. (2006). Semantic associations in Business English: A corpus-based analysis, *English for Specific Purposes, 25*(2), 217-234.
NKI (2018). *Advanced EU English course description*, National University of Public Service. Retrieved 15 March, 2018 from http://archiv.akk.uni-nke.hu/uploads/media_items/angol-szaknyelv-europai-unios-halado-angol-szaknyelv.original.pdf
Pardillos, C., Ángel, M. (Ed.). (2016). *Handbook English for judicial cooperation in civil matters*. European Judicial Training Network. Retrieved 15 April, 2019 from http://www.ejtn.eu/PageFiles/16034/Handbook_Manuel_Linguistics_Civil.pdf
Ruiz-Garrido, M. F., Palmer, J. C., Fortanet-Gómez, I., & Fortanet, I. (Eds.). (2010). *English for Professional and Academic Purposes*. Amsterdam and New York: Rodopi.
Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.
Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Scott, M. (2008). *Oxford Wordsmith Tools version 5.0*. Oxford: Oxford University Press.
Stubbs, M. (2001). *Words and phrases*. Oxford: Blackwell.
Tangpijaikul, M. (2014). Preparing business vocabulary for the ESP classroom. *RELC Journal, 45*(1), 51-65.
TEMPUS (2020). EU English training. Retrieved 15 January, 2020 from https://tka.hu/kepzesek/99/eu-english---europai-unios-ismeretek-angol-nyelven
Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
Tongpoon-Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes, 51*, 45-54.
Trebits, A. (2008). English lexis in the documents of the EU – A corpus-based exploratory study. Eotvos Lorand University: *Working Papers in Language Pedagogy, 2*, 38-54, Retrieved 15 January, 2020 from http://langped.elte.hu/WoPaLParticles/W2Trebits.pdf
Trebits, A. (2009a). Conjunctive cohesion in English language EU documents–A corpus-based analysis and its implications. *English for Specific Purposes 28*(3), 199-210.
Trebits, A. (2009b). The most frequent phrasal verbs in English language EU documents A corpus-based analysis and its implications. *System 37*(3), 470-481.

Trebits, A., & Fischer, M. (2010). *EU English: Using English in EU Contexts.* Stuttgart: Klett Verlag.
Truchot, C. (2002). *Key aspects of the use of English in Europe*. Strasbourg: Council of Europe.
Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes, 30*(2), 101-112.
West, M. (1953). *A General Service List of English words*. London: Longman.
Yang, M. (2015). A nursing academic word list. *English for Specific Purposes 37*, 27-38.
Ypsilandis, G. & Kantaridou, Z. (2007). English for Academic Purposes: Case studies in Europe, *Revista de Lingüística y Lenguas Aplicadas, 2*(1), 69-83.

Appendix

EU Wordlist

ABSENCE, ACCESSION, ACCOMPANY, ACCORDANCE, ACHIEVE, ACQUIS, ACQUISITION, ADEQUATE, ADJUSTMENT, ADMINISTRATION, ADOPT, ADVERSE, AGENCY, AGENDA, AGRICULTURE, AIM, ALIGN, ALLOCATE, ALTERNATIVE, AMEND, ANALYSE, ANNEX, ANNUAL, APPROPRIATE, APPROPRIATION, APPROVE, ASPECT, ASSESS, ASSIGN, ATTAIN, AUDIT, AUTHORISE, AUTHORITY, AWARD, BARRIER, BEHALF, BENEFICIARY, BILATERAL, BREACH, BUDGET, BURDEN, CAMPAIGN, CANDIDATE, CAPACITY, CATEGORY, CERTIFICATE, CHARTER, CIRCULATION, CIVIL, CLARIFY, CLAUSE, CLIMATE, CODE, COFINANCE, COFUND, COHERENT, COHESION, COMBAT, COMMISSION, COMMUNICATE, COMPATIBLE, COMPENSATION, COMPETENCE, COMPETITIVENESS, COMPLEMENT, COMPLEX, COMPLIANCE, COMPONENT, COMPREHENSIVE, COMPRISE, COMPULSORY, CONCEPT, CONCLUDE, CONCLUSION, CONCRETE, CONDUCT, CONFERENCE, CONFER, CONFIDENTIAL, CONFIRM, CONFLICT, CONFORMITY, CONSEQUENCE, CONSEQUENTLY, CONSIDERABLE, CONSIST, CONSISTENCY, CONSOLIDATE, CONSTITUTE, CONSTITUTION, CONSTRUCTION, CONSULT, CONSUMPTION, CONTEXT, CONTRACTUAL, CONTRARY, CONVENTION, CONVERGENCE, COOPERATE, COORDINATE, CORE, CORRESPOND, CRISIS, CRITERION, CROSS-BORDER, CRUCIAL, CURRENCY, CUSTOMS, CYCLE, DATA, DATABASE, DEADLINE, DEBATE, DECISION-MAKING, DECLARATION, DEEM, DEFINE, DELEGATE, DEMOCRACY, DEMONSTRATE, DERIVE, DESIGNATE, DIALOGUE, DIMENSION, DISABLED, DISCRIMINATION, DISPOSAL, DISPUTE, DISSEMINATE, DISTINCTION, DISTRIBUTION, DIVERSE, DOCUMENT, DOMESTIC, DRAFT, DURATION, EFFECTIVE, EFFICIENT, ELECTRONIC, ELEMENT, ELIGIBLE, ELIMINATE, EMERGE, EMPHASIS, ENABLE, ENFORCEMENT, ENHANCE, ENLARGEMENT, ENSURE, ENTAIL, ENTERPRISE, ENTITLE, ENTITY, ENTRY, ENVISAGE, EQUIVALENT, ESSENTIAL, ESTIMATE, EURO, EVALUATE, EXCEED, EXCLUDE, EXECUTIVE, EXEMPT, EXPENDITURE, EXPERTISE, EXPLOIT, EXPORT, EXTERNAL, FACILITATE, FACILITY, FACTOR, FISCAL, FLEXIBLE, FOCUS, FOLLOW-UP, FORESEEN, FORMAT, FORUM, FOSTER, FRAMEWORK, FRAUD, FULFIL, FUNCTION,

FUNDAMENTAL, FURTHERMORE, GENDER, GENERATE, GEOGRAPHICAL, GLOBAL, GOODS, GOVERNANCE, GUIDELINES, HARMONISE, HEREBY, HEREINAFTER, HERITAGE, HIGHLIGHT, IMPACT, IMPLEMENT, IMPORT, IMPORTANCE, IMPOSE, INCENTIVE, INCLUSION, INCORPORATE, INCUR, INDEPENDENCE, INFRASTRUCTURE, INITIAL, INITIATE, INNOVATION, INSTITUTE, INTEGRATE, INTERIM, INTERNAL, INTERNATIONAL, INTEROPERABILITY, INTERPRETATION, INTERVENE, INVEST, JOURNAL, JUDICIAL, JURISDICTION, JUSTIFY, LAUNCH, LEGISLATION, LEGITIMATE, LIABLE, LINK, LONG-TERM, MANDATE, MANDATORY, MANUFACTURE, MARITIME, MECHANISM, MEDIUM, METHOD, MIGRATION, MINIMUM, MOBILE, MONETARY, MONITOR, MULTILATERAL, MUTUAL, NEGATIVE, NEGOTIATION, NETWORK, NEVERTHELESS, NOTIFY, OBJECTIVE, OBLIGATION, OBTAIN, ONGOING, OUTCOME, OVERALL, PARAGRAPH, PARLIAMENT, PARTICIPATE, PENALTY, PERSONNEL, PERSPECTIVE, PHASE, PLATFORM, POTENTIAL, PRECEDE, PRECISE, PRELIMINARY, PREMISES, PRESIDENCY, PRIMARY, PRINCIPAL, PRINCIPLE, PRIOR, PRIORITY, PROCEED, PROCUREMENT, PROMOTE, PROPORTION, PROTOCOL, PROVISION, PROVISIONAL, PUBLICATION, PURSUE, RAPPORTEUR, REFORM, REGIME, REGULATION, REINFORCE, REJECT, RELEVANT, REPEAL, REQUEST, RESOLUTION, RESOURCE, RESPECTIVELY, RESPOND, RESTRICT, RETAIN, REVENUE, REVIEW, REVISION, RURAL, SAFEGUARD, SCHEME, SCOPE, SECRETARIAT, SECTOR, SIGNIFICANT, SOLE, SOLIDARITY, SOURCE, SPECIFY, STAKEHOLDER, STATISTICS, STATUS, STATUTORY, STIMULATE, STRATEGY, STRUCTURE, SUBMISSION, SUBSEQUENT, SUBSIDIARY, SUFFICIENT, SUM, SUMMARY, SUPERVISION, SURVEILLANCE, SURVEY, SUSPEND, SUSTAINABLE, TARGET, TASK, TECHNICAL, TEMPORARY, TERRITORY, TEXT, THEMATIC, THEREOF, TRANSACTION, TRANSITION, TRANSMIT, TRANSPARENCY, TRANSPORT, TREATY, TREND, UNDERTAKE, UNIFORM, UPDATE, URGENT, VALID, VERIFY, VIA, VOCATIONAL, VOLUME, VOLUNTARY, VULNERABLE, WEBSITE, WITHDRAW, WORLDWIDE
Abbreviations
DG, EC, EEA, EEC, EU, EUR, EURATOM, GDP, ICT, OJ, OLAF, SME, UN, VAT
Geographical word families
AUSTRIA, BELGIUM, BRUSSELS, BULGARIA, CYPRUS, CZECH, DENMARK, ESTONIA, EUROPE, FINLAND, FRANCE, GERMANY, GREECE, HUNGARY, IRELAND, ITALY, LATVIA, LISBON, LITHUANIA, LUXEMBOURG, MALTA, NETHERLANDS, NORWAY, POLAND, PORTUGAL, ROMANIA, SLOVAKIA, SLOVENIA, SPAIN, SWEDEN, UK