

“IN THE LIGHT OF”: A CORPUS-BASED ANALYSIS OF LEXICAL BUNDLES IN TWO EU-RELATED REGISTERS

Réka Jablonkai

Corvinus University of Budapest
rjr@uni-corvinus.hu

Abstract: This paper is a report on a corpus-based analysis of two EU-related registers. The study aims to explore the discourse of official EU texts and online EU news applying a frequency-based approach suggested by Biber and Conrad (1999). The concept developed and applied for this approach is the lexical bundle which has been found characteristic for specific registers. Based on the investigation of the most frequent 50 single-word items and the structural and functional analysis of the identified lexical bundles in official EU texts and online EU news texts, the discourse of the two EU-related registers are found to differ considerably. Discourse in official EU texts can be characterised in contrast to online EU news texts by a higher proportion of content words in the top 50 single-word lexical items, more frequent use of lexical bundles in general, very few identical lexical bundles and by a significantly different distribution of lexical bundles across structural types and functional categories. Besides the description of the two EU-related registers, pedagogical implications are also discussed.

Keywords: register analysis, lexical bundles, corpus-based analysis, EU discourse, ESP

1 Introduction

The present study is very strongly motivated by the current state of ESP teaching in higher education in Hungary. Since Hungary's accession to the EU there has been a strong demand for EU experts with a good working knowledge of English used in an EU context, for example, working at an EU institution, writing proposals for EU funds, reading relevant EU legal texts, etc. Although there are a few published materials available for teaching English for the EU offering a selection of EU texts (Lénárt, 2003), many courses are based on newspaper articles on EU issues from different sources. The main aim of the present study is to investigate to what extent the use of newspaper articles for ESP teaching purposes for an EU context is relevant and sufficient. Therefore, as a first step the linguistic variation of English used in the two contexts is compared from a lexical and discourse point of view.

1.1 Lexical bundles

Several recent studies emphasise the importance of multi-word items in spoken and written discourse (Biber, Conrad, & Cortes, 2004; Cowie, 1992; Moon, 1998; Moon, 2000). Analysing recurrent word combinations or longer lexical phrases, clusters, chunks or idioms has been an especially productive approach in describing discourse in certain contexts like university registers, academic registers (Biber, & Conrad, 1999; Biber, & Barbieri, 2007; Hyland, 2008) or the language of journalistic prose (Cowie, 1992). These multi-word items have been studied

under different terms such as fixed expressions (Moon, 1998; Moon, 2000), lexical phrases (Nattinger, & DeCarrico, 1992), pre-fabs, ready made units (Cowie, 1992), using different criteria to define and identify multi-word items and thus throwing light on different aspects of structures and functions of multi-word items in discourse. Earlier studies of longer word combinations were based on perceptual salience rather than empirical evidence in real language use (Biber, & Conrad, 1999). With the advent of computer technology and corpus linguistics it became possible to investigate longer sequences of words in discourse statistically, that is, researchers could focus on what is frequent instead of examining what stands out. The concept of lexical bundles was introduced in *The Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Biber, *et al.* (1999) distinguished lexical bundles from both collocations and idioms. According to their distinction, idioms are the most idiomatic and invariable type of multi-word items and are usually structurally complete units. Collocations, on the other hand, are statistical associations between two words that are variable and not idiomatic in the sense that in a collocation words can be associated with several other words and they retain their own meaning. Biber and Conrad (1999) defined lexical bundles as 'the most frequent recurring lexical sequences; which can be regarded as extended collocations: sequences of three or more words that show a statistical tendency to co-occur (e.g., *in the case of the*)' (p.183). The concept of lexical bundles has been used in several later studies (Biber, *et al.*, 2004; Biber, & Conrad, 1999; Biber, & Barbieri, 2007; Cortes, 2004; Hyland, 2008) to investigate common multi-word items in discourse focusing mainly on registers in university and academic contexts.

1.2 Functions and structures of lexical bundles

One of the early findings about lexical bundles was that they are present in written and spoken registers alike and they were considered 'basic building blocks for constructing spoken and written discourse' (Biber, & Conrad, 1999, p.188). Moreover, further research found that in certain written registers like written course management 'comprising of syllabi and description of course assignments' lexical bundles are surprisingly common (Biber, & Barbieri, 2007). These results were in contrast to previous analyses which regarded the use of prefabs a characteristic of spoken registers (Pawley, & Syder, 1983).

As regards the structure of lexical bundles, previous research found that lexical bundles are structurally complex, usually incomplete and not fixed (Biber, & Conrad, 1999). Comparing bundles across registers showed that the grammatical structure of lexical bundles is a distinct characteristic of registers. Biber and Conrad (1999) reported that the most frequent lexical bundles in conversation have the pattern personal pronoun + verb phrase (clause-fragment), for example, *I don't know how, you might as well*, whereas in academic prose the two most important patterns are noun phrase and prepositional phrase fragments, for example, *one of the most, an increase in the*. Structural types of lexical bundles were further analysed and classified by Biber, *et al.* (2004). Their framework, which was applied in the present study, is described in detail in Section 2.

Previous research has also looked into the discourse functions of lexical bundles. By developing a detailed taxonomy, Biber, *et al.* (2004) found that the three main functions lexical

bundles serve in discourse include expressing stance, organising discourse and referring to, for example, specific attributes, time and place. The taxonomy is outlined in Section 2.

1.3 Multi-word items in language teaching

Research on multi-word items has already emphasised the important role of longer lexical sequences in language teaching (Cowie, 1992). Still there are only a few studies that focus on the role of lexical bundles in language teaching (Cortes, 2004; Cortes, 2006; Scott, & Tribble, 2006). Most of these investigations compare student writing with expert writing in academic disciplines. Scott and Tribble (2003), by looking at apprentice and expert literary papers conclude that such a comparison can be helpful for students to find what they lack to become proficient writers. Cortes (2004), by looking at texts in two disciplines, namely, history and biology, also compared the use of lexical bundles in student and expert writing and found that students did not often use lexical bundles or used them in a different way. In her later study she concludes that having a few lessons with a few examples of lexical bundles in expert writing will not result in students using more lexical bundles in a more appropriate way, but increases their interest in and awareness of these expressions (Cortes, 2006).

The present study extends the research on lexical bundles by investigating registers outside the university and academic contexts. The research framework of lexical bundles was used to analyse and compare two EU-related registers. The two registers can be defined as follows: the register of online EU news texts comprises news texts on EU issues in online news sources and the register of official EU texts can be found in texts produced by EU institutions. Frequent lexical bundles were identified and compared in order to draw conclusions on the appropriateness and usefulness of these EU-related registers for ESP teaching purposes. The starting point for this study was the assumption that lexical bundles are register-specific, that is, each register employs a distinct set of lexical bundles, associated with the typical communicative purposes of that register (Biber, & Barbieri, 2007, p.265) which has been demonstrated by findings of previous research (Biber, & Conrad, 1999; Hyland, 2008; Scott, & Tribble, 2006).

Therefore, the present study is motivated by the following broad research question: What differences are there between the discourse of English texts produced by EU institutions and the discourse of online EU news texts on EU issues from an ESP pedagogic perspective?

The paper is organised in the following way: Section 2 outlines the methodological issues of the study describing the Online EU news corpus and the EU discourse corpus, and the methods and tools to analyse the corpora and to identify lexical bundles. It also provides the taxonomies applied in the structural and functional analysis of the lexical bundles identified. Section 3 reports on the main findings, comparing two lexical aspects of the EU-related corpora, namely, most frequent single-word items and the lexical bundles with similar findings in other registers. Section 4 discusses conclusions and pedagogical implications from an ESP perspective.

2 Method

Two corpora were created for the analysis, one comprising newspaper articles dealing with EU issues from several English online EU news (Online EU news corpus) and another consisting of official EU documents like Commission and Council press releases, legal texts from the Official Journal of the European Union (EU discourse corpus). Both corpora are around 120,000 running words. The detailed contents of the corpora are outlined in Table 1 and 2. Texts for the corpora were downloaded from the Internet and were selected randomly from the time period of January to August 2007.

Online EU news source	Number of texts	Number of words
BBC	8	8,834
Financial Times	10	8,020
European Voice	55	33,408
Guardian	22	16,334
International Herald Tribune	36	21,946
Times Online	21	16,000
USA Today	27	17,561
Total:	179	122,103

Table 1. English Online EU news corpus

EU texts	Number of texts	Number of words
Legal texts of the EU	35	56,888
EU Press releases	28	62,414
Total:	63	119,302

Table 2. English EU discourse corpus

The analysis of the corpora had three stages. First, the software AntConc was used to create frequency lists of single-word lexical items and four-word lexical items. The function Word List was used for the first task, the function Clusters with n-grams set at 4 was used for the second task. In both cases the tool was set to analyse all data in lower case. When creating the frequency list, hyponyms were checked and cleared for the first 50 items. Although Scott and Tribble (2006) argue that three and four-word lexical items are both good discriminators of registers and three-word lexical items have advantages especially for pedagogic purposes, the present study focused on four-word lexical bundles for two reasons. One is that most studies on lexical bundles analyse bundles with four words and therefore lexical bundles in EU-related registers can be compared to bundles in other registers like university or academic registers (Biber, *et al.*, 2004; Biber, & Barbieri, 2007; Cortes, 2004; Hyland, 2008). Furthermore, as Cortes (2004) also argues three-word lexical bundles are often part of four-word bundles and four-word bundles are more frequent and give more variety for the structural and functional analysis than five-word bundles.

At the second stage of the investigation lexical bundles were selected from the automatically identified four-word lexical items. Lexical bundles were defined based on Biber and Conrad (1999) as sequences of three or more words that show a statistical tendency to co-occur (p.183). In order to be considered lexical bundles, the four-word lexical items must recur at least 40 times per million (Biber, & Barbieri, 2007). As the present study used small corpora, normalised rates of occurrences had to be used. Therefore the following requirements were applied to identify lexical bundles in the present study. Taking a fairly conservative approach the necessary frequency was set at 5 in the whole corpus which represents the 40 per million occurrence. In order to avoid the impact of idiosyncratic use lexical bundles were defined, in addition to their overall frequency in the corpora, on their distribution in sub-corpora and individual texts. Therefore, the following requirement was introduced: only recurring four-word lexical items occurring in at least two sub-corpora and five different texts (Biber, & Barbieri, 2007; Cortes, 2006) can be included in the analysis. Setting the necessary number of occurrences in individual texts is a common defining characteristic of lexical bundles. In the current study the additional requirement was introduced in the case of the online EU news texts because several newspapers have the same journalists who report on certain issues, as a consequence, controlling only for individual texts might not be sufficient to avoid idiosyncratic language use. Therefore bundles that occurred at least 5 times in the corpus in 5 different texts and in at least 2 different news sources were included in the analysis. In the case of the official EU texts the requirement that the four-word lexical item should occur in both the legal texts and the press releases was introduced to ensure that lexical bundles represent EU discourse rather than the language used in EU press releases or legal texts.

The qualitative analysis of the identified lexical bundles in the two EU-related registers included their structural and functional analysis. The structural analysis applied the structural types of lexical bundles identified by Biber, *et al.* (2004). According to this classification, there are three main structural types which include 1) lexical bundles that incorporate verb phrase fragments like *that's one of the, is based on the*, 2) lexical bundles that incorporate dependent clause fragments like *that this is a, to come up with*, 3) lexical bundles that incorporate noun phrase and prepositional phrase fragments like *at the end of, at the same time*. Each main structural type entails several sub-types as illustrated in Table 3.

Structural types	Sub-types	Sample bundles
1. Lexical bundles that incorporate verb phrase fragments	1.a 1st/2nd person pronoun + VP fragment	<i>I'm not going to</i>
	1.b 3rd person pronoun + VP fragment	<i>and this is a</i>
	1.c discourse marker + VP fragment	<i>I mean I don't</i>
	1.d Verb phrase (with non-passive verb)	<i>have a lot of</i>
	1.e Verb phrase (with passive verb)	<i>is based on</i>
	1.f yes-no question fragments	<i>are you going to</i>
	1.g WH- question fragments	<i>what do you think</i>
2. Lexical bundles that incorporate dependent clause fragments	2.a 1st/2nd person pronoun + dependent clause fragment	<i>I want you to</i>
	2.b WH-clause fragments	<i>when we get to</i>
	2.c If-clause fragments	<i>if we look at</i>
	2.d to-clause fragment	<i>to be able to</i>
	2.e That-clause fragment	<i>that this is a</i>
3. Lexical bundles that incorporate noun phrase and prepositional	3.a Noun phrase with of-phrase fragment	<i>one of the things</i>

phrase fragments	3.b Noun phrase with other post-modifier fragment	<i>the way in which</i>
	3.c Other noun phrase expressions	<i>a little bit more</i>
	3.d Prepositional phrase expressions	<i>at the end of</i>
	3.e Comparative expressions	<i>as well as the</i>

Table 3. Structural types of lexical bundles (Biber, *et al.*, 2004, p.381.)

The functional analysis applied the taxonomy of discourse functions of lexical bundles outlined by Biber, *et al.* (2004). According to their taxonomy, lexical bundles serve three main discourse functions in registers: 1) stance bundles express attitude or assessment, 2) discourse organisers reflect the relationships between different parts of texts, 3) referential expressions refer to physical or abstract entities, or to other textual parts. Each of these main categories has several sub-categories which are associated with more specific discourse functions. The main discourse function categories with their sub-categories, as developed by Biber, *et al.* (2004), are shown in Table 4.

Categories	Sub-categories	Sample bundles
I. Stance bundles	A. Epistemic stance	<i>the fact that the, and I think that</i>
	B. Attitudinal/ modality stance	
	B1) Desire	<i>what do you want</i>
	B2) Obligation/ directive	<i>it is important to</i>
	B3) Intention/ Prediction	<i>it's going to be</i>
	B4) Ability	<i>it is possible to</i>
II. Discourse organisers	A. Topic introduction	<i>in this chapter we</i>
	B. Topic elaboration/ clarification	<i>on the other hand</i>
III. Referential bundles	A. Identification/ focus	<i>is one of the</i>
	B. Imprecision	<i>or something like that</i>
	C. Specification of attributes	
	C1) Quantity specification	<i>a lot of people</i>
	C2) Tangible framing	<i>in the form of</i>
	C3) Intangible framing	<i>on the basis of</i>
	D. Time/ Place/ Text reference	
	D1) Place reference	<i>in the United States</i>
	D2) Time reference	<i>at the same time</i>
	D3) Text-deixis	<i>as shown in figure</i>
D4) Multi-functional reference	<i>in the middle of</i>	

Table 4. Discourse functions of lexical bundles (based on Biber, *et al.*, 2004, pp.384-388)

The Concordance function of AntConc was used for the analysis of the discourse functions of the lexical bundles in the two registers. The software provided the context by concordancing and the functions the lexical bundles in question performed were analysed manually.

Finally, the structural and functional distribution of lexical bundles in the two EU-related registers was compared. The comparison included statistical tests that were carried out with the help of the SPSS statistical software. The applied statistical tests were the log-likelihood chi-square (LL) test for the structural distribution and Fisher's exact test for the functional distribution of lexical bundles. The reasons behind the choice of tests were the following: (1) McEnery, Xiao, and Tono (2006) prefer the LL test to the simple chi-square test in the case of corpus data, as it does not assume that data are normally distributed which is the case with samples of natural language; (2) in the case of the functional distribution because of the small sample size the number of lexical bundles for certain types, for example, Stance bundles, did not reach the necessary level of 5 to make chi-square or LL tests reliable. Therefore, the more reliable Fisher's exact test was applied as suggested by McEnery, *et al.* (2006).

The following more specific research questions helped to focus the study:

- a. What single-word lexical items are especially associated with EU discourse and online EU news on EU issues?
- b. What lexical bundles occur in EU discourse?
- c. What lexical bundles occur in online EU news on EU issues?
- d. How do lexical bundles identified in EU discourse differ from lexical bundles in online EU news?
- e. What are the pedagogical implications of the findings for EU English courses?

3 Lexical analysis of EU-related registers

In this section the findings of the lexical analysis of the two EU-related registers are commented on in the following order: first, the most frequent single-word items are compared across the two EU-related corpora and a general corpus, second the four-word lexical bundles identified in the corpora are compared to each other and to lexical bundles in other registers. Analysis of lexical bundles includes investigating their overall frequency, their structure and discourse function and statistical tests to demonstrate the significance of the differences between the two corpora.

3.1 Frequency lists compared

The frequency-ordered word list of a corpus shows the word types that the corpus contains ranked according to the number of occurrences of each word type. Comparing frequency lists of single-word lexical items can give interesting insights into the differences between the texts in each corpus (Hunston, 2002; Scott, & Tribble, 2006). As shown in Table 5, the first four items in each of the frequency lists are the same across registers. This is, however, not surprising as in most corpora these function words occupy the first few places in frequency lists. An aspect worth investigating is the proportion of content words among the most frequent lexical items. In the case of specialised corpora there is usually a higher proportion of content words among the most frequent items. This is also what we find in the Online EU news and EU discourse corpora. These content words like *EU*, *Commission*, *European* are related to EU issues in general. The only content word in the general written corpus is *said* which can also be found in the Online EU news corpus.

Rank	Most frequent words in Online EU news			Most frequent words in EU discourse			Most frequent words in a general written corpus sample from CIC		
	Freq.	Word	%	Freq.	Word	%	Freq.	Word	%
1	8504	the	6.96	10142	the	8.50	284174	the	5.68
2	4054	to	3.32	5589	of	4.68	132335	to	2.65
3	3498	of	2.86	4071	and	3.41	125526	and	2.51
4	2769	and	2.27	3184	to	2.67	122903	of	2.46
5	2692	a	2.20	2684	in	2.25	114381	a	2.29
6	2447	in	2.00	2023	for	1.70	84940	in	1.70
7	1668	that	1.37	2001	a	1.68	59454	was	1.19
8	1400	EU	1.15	1355	on	1.14	51642	it	1.03
9	1285	for	1.05	1075	be	0.90	50871	I	1.02
10	1268	is	1.04	957	by	0.80	50007	he	1.00
11	1245	on	1.02	858	that	0.72	46195	that	0.92
12	959	it	0.79	848	with	0.71	41607	she	0.83
13	892	said	0.73	798	is	0.67	41606	for	0.83
14	856	be	0.70	795	as	0.67	38361	on	0.77
15	841	by	0.69	778	Council	0.65	36500	her	0.73
16	725	with	0.59	759	European	0.64	35773	you	0.72
17	691	European	0.57	675	Commission	0.57	34871	is	0.70
18	672	as	0.55	549	at	0.46	33829	with	0.68
19	623	are	0.51	530	this	0.44	32535	his	0.65
20	621	have	0.51	521	Member	0.44	31420	had	0.63
21	610	will	0.50	450	or	0.38	30993	as	0.62
22	594	from	0.49	445	Minister	0.37	29026	at	0.58
23	573	not	0.47	438	regulation	0.37	26134	but	0.52
24	557	would	0.46	430	EU	0.36	26122	be	0.52
25	556	at	0.46	419	which	0.35	22805	have	0.46
26	551	but	0.45	412	from	0.35	21574	from	0.43
27	533	this	0.44	406	are	0.34	21554	not	0.43
28	514	was	0.42	393	an	0.33	21097	they	0.42
29	486	has	0.40	386	shall	0.32	20391	by	0.41
30	452	its	0.37	374	it	0.31	17577	this	0.35
31	440	an	0.36	360	its	0.30	17227	are	0.34
32	416	he	0.34	351	L	0.29	16363	were	0.33
33	416	we	0.34	340	not	0.28	16240	all	0.32
34	408	which	0.33	325	States	0.27	15647	him	0.31
35	387	new	0.32	309	C	0.26	15526	up	0.31
36	369	more	0.30	304	Mr	0.25	15431	an	0.31

37	367	Europe	0.30	300	will	0.25	15255	said	0.31
38	352	Commission	0.29	286	should	0.24	14913	there	0.30
39	343	they	0.28	258	decision	0.22	14525	one	0.29
40	321	states	0.26	256	EC	0.21	14493	been	0.29
41	317	their	0.26	256	government	0.21	14445	would	0.29
42	297	all	0.24	247	Union	0.21	14337	out	0.29
43	290	year	0.24	234	EN	0.20	13804	so	0.28
44	288	or	0.24	233	has	0.20	13788	their	0.28
45	288	there	0.24	232	such	0.19	13646	what	0.27
46	268	member	0.22	228	Community	0.19	13566	when	0.27
47	260	US	0.21	225	article	0.19	13526	we	0.27
48	253	up	0.21	224	other	0.19	13313	if	0.27
49	252	also	0.21	220	State	0.18	13035	me	0.26
50	251	countries	0.21	212	energy	0.18	12930	my	0.26

Table 5. The 50 most frequent words in three corpora
(source of written general corpus: O'Keefe, McCarthy, & Carter, 2007, p.35-36)

Further analysis of the content words in the two EU-related registers show that most of the content words are nouns, like *Europe*, *Commission*, *states*, *year*, *member*, *countries* in the Online EU news corpus and *Council*, *Commission*, *Member*, *Minister*, *regulation*, *States*, *decision*, *government*, *Union*, *Community*, *article*, *State*, *energy* in the EU discourse corpus. As shown in Table 6, there are, however, more than twice as many nouns in the top 50 words in the EU discourse corpus than in the Online EU news corpus. Another characteristic feature of the specialised corpora is the common use of abbreviations such as *EU*, *US* in the Online EU news corpus and *EU*, *L*, *C*, *Mr*, *EC*, *EN* in the EU discourse corpus. There are also more abbreviations among the most frequent words in the EU discourse corpus. The higher proportion of content words and especially nouns and abbreviations in the top 50 words indicate a more technical type of discourse in the official EU text collection.

	Online EU news	EU discourse	general written corpus
verb	1	0	1
noun	6	13	0
adjective	2	1	0
abbreviation	2	6	0

Table 6. Categories of content words among the 50 most frequent lexical items

Although the frequency of single-word items threw light on some differences in the lexical characteristics of the EU-related registers, further analysis was needed in order to gain deeper insights into how these lexical items co-occur and what frequent word combinations they form. These frequent word combinations are discussed and compared in the following section.

3.2 Lexical bundles in EU-related registers

As shown in Table 7, the range of lexical bundles is fairly similar in the two registers. However, if the proportion of bundles in the texts is compared, there is a considerable difference. The percentage of words in bundles is more than twice as high in the EU discourse corpus. The reason for this might be that there are certain pre-formulated multi-word items which are very frequently used in EU texts reflecting a high level of formality and leaving little room for creativity. The most common forty bundles in the EU discourse corpus occur more than 200 times per million, the first two, *of the European Union* and *official journal of the*, being as frequent as 1526 and 1467 per million respectively. These two are, however, outstanding as the third and fourth bundles in the frequency list *on the basis of* and *in accordance with the* have a frequency around 300 per million and the rest of the lexical bundles are more evenly distributed. By contrast, the most frequent ten lexical bundles in the Online EU news corpus, *at the end of*, *# per cent to*, *the EU and the*, *German chancellor Angela Merkel*, *the end of the*, *of the European Union*, *the council of ministers*, *at the same time*, *presidency of the EU*, *the Treaty of Rome*, have a frequency between 80 to 164 per million and all the bundles are distributed fairly evenly.

Corpus	Different bundles	Total cases	% of total words in bundles
Online EU news	83	632	2
EU discourse	89	1373	4.6

Table 7. Bundle frequency information

The comparison of lexical bundles across registers showed that there is a considerable difference in the bundles themselves. Table 8 illustrates this difference by highlighting the bundles that are shared with the top 40 lexical bundles in three registers with the BNC_ALL as a reference. Considering that lexical bundles are ‘building blocks of discourse’, the less the shared area, the greater the contrast between the discourse of the registers represented by the corpora. Results indicate that there is only a slight difference between the two EU-related registers when compared to a general corpus. Findings highlight also that these two registers differ considerably from academic prose in general.

In contrast to a general corpus, the two EU-related registers show similarity. Comparing all the lexical bundles in the two EU-related registers, however, shows important differences. This comparison was made considering several aspects. First, the lexical bundles themselves were compared, then their grammatical structure was analysed, and finally, their discourse function in context was investigated.

The first step resulted in nine shared lexical bundles in the two registers. These are the following: *at the end of*, ***of the European Union***, *at the same time*, ***in the light of***, *on the basis of*, *as part of the*, ***justice and home affairs***, *at the time of*, ***of the European Parliament***. Out of these nine lexical bundles five can also be found in the top 40 of the general corpus which suggests that these are more general lexical bundles that are frequently used in several registers. The other four bundles ó highlighted in bold above ó seem to be specific for these two EU-related registers. The fact that the proportion of the same topic-related lexical bundles in all the bundles is fairly small, 4.5% in the EU discourse corpus and 4.8% in the Online EU news corpus, indicates considerable difference between the general corpus and the two EU-related corpora.

BNC ALL	BNC_ACAD	Online EU news	EU discourse
the end of the	as a result of	the end of the	at the end of
at the end of	as well as the	at the end of	at the same time
at the same time	at the end of	at the same time	on the basis of
for the first time	at the same time	for the first time	at the time of
on the other hand	at the time of	By the end of	as well as the
as a result of	in terms of the	On the basis of	in the form of
the rest of the	in the case of	# per cent to	in the context of
per cent of the	in the context of	the EU and the	as part of the
in the case of	in the form of	German Chancellor Angela Merkel	of the European Union
one of the most	it is possible to	of the European Union	official journal of the
the secretary of state	on the basis of	the council of ministers	in accordance with the
by the end of	on the other hand	presidency of the EU	of the member states
is one of the	one of the most	the Treaty of Rome	minister of state for
to be able to	per cent of the	EU and the US	set out in the
on the basis of	the end of the	and the Czech Republic	as last amended by
I'm going to	the fact that the	between the EU and	member states and the
in the middle of	the nature of the	French and Dutch voters	of the European Parliament
at the time of	the rest of the	in the council of	the European Parliament and
as well as the	the way in which	president of the European	and of the Council
in the form of	a wide range of	by French and Dutch	be sent to the
the top of the	as shown in fig	charter of fundamental rights	European Parliament and of
I think it's	as we have seen	commission president Jose Manuel	as set out in
a member of the	can be used to	in France and the	parliament and of the
the way in which	in relation to the	in the light of	the member states and
the fact that the	in the absence of	president Jose Manuel Barroso	in the light of
you've got to	in the course of	Prime Minister Tony Blair	the basis of a
a great deal of	in this case the	the European Commission is	in line with the
was one of the	is likely to be	the European court of	the basis of the
a wide range of	it is clear that	the European Union and	within the scope of
the middle of the	it is difficult to	the European Union on	in addition to the
the nature of the	it is important to	at a time when	justice and home affairs
at the beginning of	it is necessary to	European Commission President Jose	on health and safety

will be able to	more likely to be	for the EU to	the implementation of the
the back of the	on the part of	France and the Netherlands	the official journal of
in terms of the	that there is a	joined the EU in	a high level of
I'd like to	that there is no	per cent in the	fall within the scope
it is possible to	the extent to which	Prime Minister Romano Prodi	having regard to the
secretary of state for	the size of the	said that the EU	in the member states
in the context of	the structure of the	50th anniversary of the	in the official journal
as part of the	the ways in which	the European Central Bank	on the protection of

Table 8. Top 40 four-word lexical bundles in three registers compared to BNC_ALL (source of BNC data: Scott, & Tribble, 2006, p.140)

3.3 Structure of lexical bundles

The next step is to compare what structural types the lexical bundles represent in the two corpora. As illustrated in Figure 1 on the following page, all three main structural types can be found in both corpora. The majority of the bundles incorporate noun phrases and prepositional phrases. This corresponds to earlier findings as written registers like textbooks or academic prose (Biber, *et al.*, 2004) were found to operate with bundles of this structural type. Having a closer look at this type of bundles, illustrated in Table 9, it may be seen that most of the noun phrases are noun phrases with *of*-phrase fragments like *presidency of the EU*, *court of first instance*, *the support of the* in the Online EU news corpus and *official journal of the*, *the council of the*, *the results of the* in the EU discourse corpus. This is the majority of bundles in both registers, which again corresponds to findings from academic registers as noun phrases with *of*-phrase fragments were found to be the most frequent in several academic disciplines (Hyland, 2008). Another similarly high proportion is taken by prepositional phrases. Their proportion is almost 10 percent higher in EU texts. As regards their discourse function, which will be dealt with in Section 3.4 in more detail, these are typical linking expressions like *in accordance with the*, *in the light of*, *in addition to the*, *with regard to* in the EU discourse corpus and *on the other hand*, *in the light of* in the Online EU news corpus.

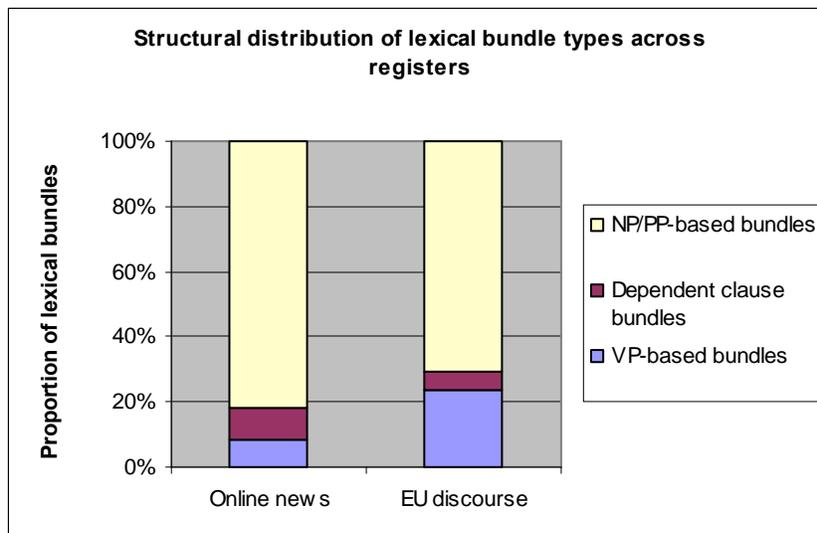


Figure 1. Structural distribution of lexical bundle types across registers

More interesting is the difference between the two registers. Besides the difference in the proportion of prepositional phrases, there is a striking difference in the other noun phrase expressions category. What can be found in this category in the online EU news register is several names, positions of people, countries relevant from a European and a more global perspective like *and the United States, German Chancellor Angela Merkel, President Jose Manuel Barroso, British Prime Minister Tony* or institutions like *the European Central Bank, the World Trade Organization*, whereas texts in the EU discourse corpus employed considerably fewer bundles of this type, referring exclusively to countries and institutions within the EU like *the member states and, the European Parliament and, the Commission and the, and the United Kingdom*.

Structural type	Online EU news	EU discourse
Verb phrase with non-passive verb	7.2	11.2
Verb phrase with passive verb	1.2	12.4
WH-clause fragment	2.4	0
<i>to</i> -clause fragment	4.8	3.4
<i>that</i> -clause fragment	2.4	2.2
Noun phrase with <i>of</i> -phrase fragment	36.1	31.5
Other noun phrase expression	24.1	7.9
Prepositional phrase expression	21.7	30.3
Comparative expression	0	1.1
Total	100	100

Table 9. Distribution of lexical bundles across structural sub-types in %

Another major difference can be seen in the proportion of bundles incorporating verb phrases. Based on earlier findings on written registers it was expected that official EU texts would also make use of fewer bundles of this type, as they were found to be the least common in textbooks and academic prose (Biber, *et al.*, 2004). Hyland (2008), however, found that Science and Engineering texts made considerable use of lexical bundles incorporating verb-phrases with

passive verbs mainly for guiding the reader through the text or identifying the basis for an argument. In the case of official EU texts, half of all verb-phrase bundles incorporate passive verb phrases. In online EU news texts there are altogether considerably fewer bundle types with verb phrases and there is only one incorporating a passive verb phrase (*rejected by French and*). In addition to this, these types of bundles are considerably less frequent in the Online EU news corpus than in the EU discourse corpus.

Structural types	Online EU news	EU discourse
1. Lexical bundles that incorporate verb phrase fragments	7	21
2. Lexical bundles that incorporate dependent clause fragments	8	5
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments	68	63
Total:	83	89

Table 10. Number of lexical bundles in the two corpora according to structural types

The statistical test comparing the number of lexical bundles of different structural types in the two corpora (Table 10) resulted in a log-likelihood ratio of 8.006 for 2 degrees of freedom, and the two-sided exact significance level is 0.018. It shows that the difference in the grammatical structures that are incorporated in the lexical bundles of the two EU-related registers is statistically significant at the $p=0.018$ level.

This statistically significant difference between the two registers might be explained by the audiences the two types of texts are written to. On the whole, texts in the EU discourse corpus seem to employ structures that are characteristic of more technical types of texts, like texts analysed by Hyland (2008), because their readership comprises mainly EU experts, whereas online EU news texts are written for the general public. It remains, however, unclear why there are so few bundles incorporating verb-phrases in the online EU news texts, although news texts are about events where usually things happen. One reason might be that these happenings are so diverse that they are expressed with several verbs rather than employing the same expressions several times.

3.4 Discourse function of lexical bundles

The taxonomy described in Biber, *et al.* (2004) was used to classify the lexical bundles according to their functions in context in both registers. Some new categories had to be created in order to classify bundles which performed functions that had not been identified in earlier research. The new main category introduced was subject-specific bundles. Lexical bundles related to the topic of the texts were put into this category. In the case of the Online EU news corpus this group accounted for almost 60 percent and in the case of the EU discourse corpus it was almost 40 percent of all bundles. As these figures suggest that this type of lexical bundle is important in the analysis of these two registers, they are discussed separately in Section 3.5.

Three sub-categories have been added to the original taxonomy. These categories were structure signal to the category discourse organisers and quality specification and intertextual to the category of referential bundles. The lexical bundle type categorised as structure signal was

found in the Online EU news corpus. Following Hylandø (2008) classification specifically designed for research writing, this category comprises bundles that organise stretches of discourse within the text. The example found in online EU news texts is used for signalling, in most cases, the end of a quote from a statement, as in the following example:

"With the necessary political commitment and follow-up on all sides, the new framework will deliver lower costs for businesses and consumers on both sides of the Atlantic," Barroso *said in a statement* Friday. (Online EU news)

Examples of the other two additional sub-categories were found in the EU discourse corpus. In the original taxonomy there is a sub-category for references within the text, for example, *as shown in figure*, in the EU text collection, however, there is also a bundle type which refers to other texts. Therefore the sub-category intertextual referential bundle was created for the following bundle type:

Having regard to the Treaty establishing the European Community, and in particular Article 211 thereof, (EU discourse)

Finally, a third additional sub-category was created for bundles that express quality attributes, therefore these were classified as a sub-category of referential specification bundles. An example for this type of bundle is given below:

To ensure *a high level of* protection of human life and health, the safety of additives, enzymes and flavouring for use in food stuffs for human consumption must be assessed before they are placed on the Community market, in the light of harmonised rules. (EU discourse)

3.4.1 Stance bundles

As shown in Figure 2, stance bundles were the fewest among all functional bundle types in both registers. Stance bundles in general give an interpretative frame for the following proposition and are more common in spoken registers (Biber, *et al.* 2004). The lexical bundles identified in the two EU related registers were found to express attitude or modality, more specifically, desire, obligation and intention in the online EU news corpus and obligation in the EU discourse corpus, as illustrated in the following examples in their original context:

But the German presidency of the EU is in favour of the plans and *would like to see* agreement on them by June. (Online EU news)

Here it seems *the EU wants to* make it formal. (Online EU news)

In this context, *it is important to* proceed with actions to increase the environmental performance of the European transport system. (EU discourse)

The difference between the two registers lies in the fact that EU texts exclusively employed bundles that express obligation, whereas online EU news texts made use of a slightly wider range of attitude bundles. This might be, on the one hand, because the function of most EU texts, especially EU legal texts, is to express what the EU in general or an EU institution in

particular requires from member states or other participants. On the other hand, news texts usually describe the several different attitudes participants in an event might have towards certain issues.

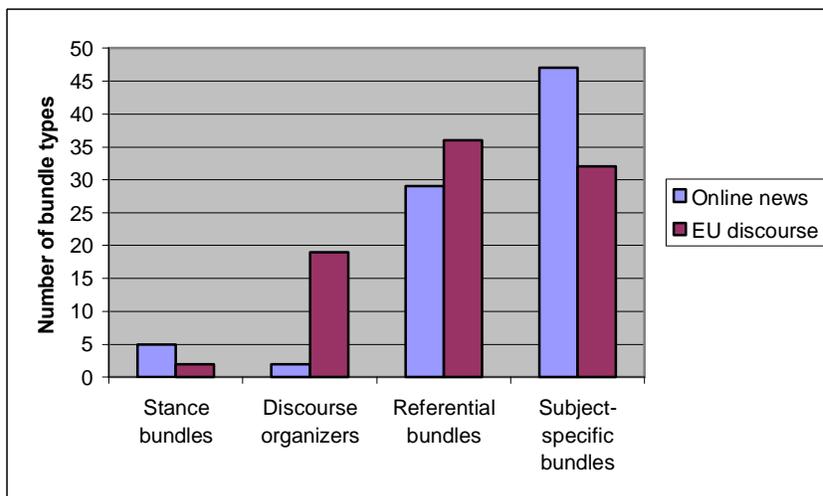


Figure 2. Functional distribution of lexical bundles across registers

3.4.2 Discourse organising bundles

The bundle category where there is a marked contrast between the two registers is that of discourse organisers. Discourse organisers perform several functions, introducing a topic, clarifying a topic, expressing contrast and comparison (Biber, *et al.* 2004, Cortes, 2004) and signalling structure. The following sentences from the corpora exemplify the use of these bundles in context:

On the other hand, some countries argue there is no urgent need for institutional reform and that the EU should concentrate instead on policies that deliver immediate practical benefits for citizens. (Online EU news)

The label shall bear the logo and wording *as set out in* the specification. (EU discourse)

Enduring non-inflationary growth *in line with the* provisions of the EC treaty needs to be secured in the Member States í (EU discourse)

In addition to the traditional criteria (universality, transparency, continuity, accessibility, etc), these characteristics refer to the organisational conditions and modalities applying to those services. (EU discourse)

As shown in Figure 2, texts in the EU discourse corpus employ almost ten times more lexical bundles of the discourse organising type than texts in the Online EU news corpus. Although further investigation is needed to clearly understand the reasons for this marked difference, one of the reasons might be the length of texts in the corpora. EU texts are considerably longer, which might explain the abundance of discourse organising bundles in this register. Another reason might be that online EU news texts might use a wider variety of

linguistic means or lexical items for signalling structure and organising discourse, therefore, they were not identified as frequently occurring word combinations in this corpus.

3.4.3 Referential bundles

Previous research suggests that overall referential bundles are one of the most frequently used lexical bundles in written registers. Biber and Barbieri (2007) found that out of the four written university registers they analysed three had referential bundles as the most common bundle type. Therefore, it is not surprising that in both written EU-related registers referential bundles are among the most frequently used lexical bundles. These are most frequent in the EU discourse corpus and second most frequent in the Online EU news corpus. This category is also divided into sub-categories and differences between the two registers become clearer when analysing sub-categories, as illustrated in Figure 3.

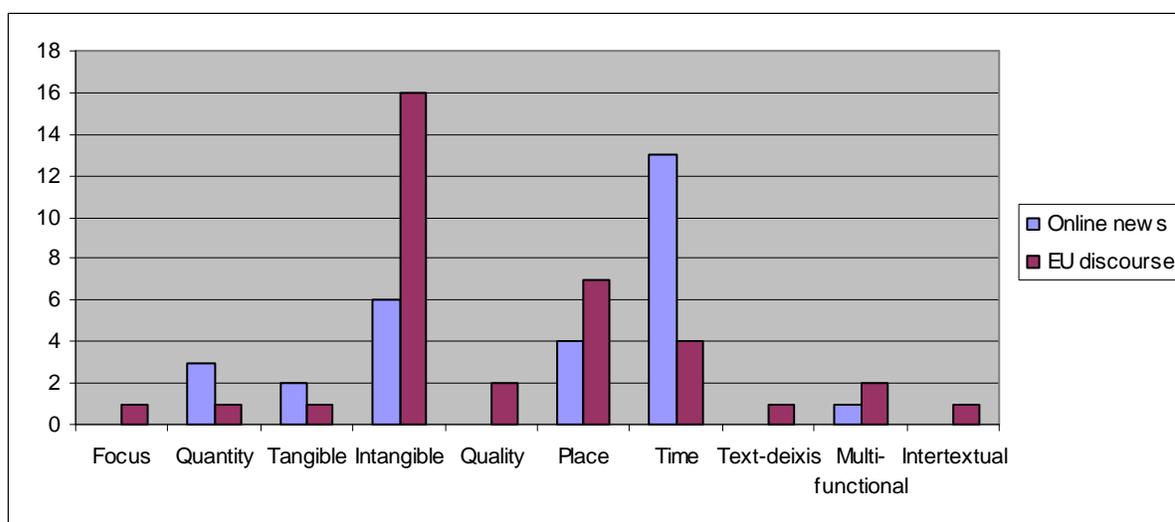


Figure 3. Referential bundles across registers

Comparing the two registers showed that more bundle types serve more functions in the EU discourse corpus. The most frequent referential bundle type in the EU discourse corpus is the intangible framing type that is the category of abstract specifying bundles used to establish logical frames and relationships in a text (Biber, et al., 2004):

On preliminary examination, the Commission finds that the notified transaction could fall *within the scope of* Regulation (EC) No 139/2004. (EU discourse)

Furthermore, more detailed reasons *in relation to the* individuals listed in that Annex should be provided. (EU discourse)

They insist on the need to preserve *the substance of the* innovations agreed upon in the 2004 IGC and to ensure as far as possible the readability and simplicity of the new Treaty. (Online EU news)

These types of bundles are considerably more frequent in EU texts than in online EU news texts, probably because news texts report on concrete events and they apply more tangible framing bundles like:

Officials suggested that the voting system could be linked in the IGC to *the size of the Commission*, an issue that French President Nicolas Sarkozy is keen to raise. (Online EU news)

Overall in EU discourse more bundle types within the referential category refer to different parts of the same text or to other texts or give the focus or logical relations within the text. Focus, intangible framing, text-deixis and intertextual bundles belong to these types as illustrated in the following examples:

This summary sets out the *main elements of the* product specification for information purposes. (EU discourse)

It shall be accompanied by the undermentioned documents and the following particulars in triplicate: 3.2.1. a description of the vehicle type *with regard to the* items mentioned in paragraph 2.2.

The most common referential bundle type in the Online EU news register specifies time. There are more than three times more bundles of this type in online EU news texts than in EU texts. The same reasoning outlined above might be applied here: concrete events have to be placed in time and their relation to each other should be clear when reporting, that is why, news texts employ more time reference bundles as in these examples:

Next Tuesday (15 May) EU leadersø negotiators on the new treaty will meet *for the first time* in Berlin to try to clarify what EU governments can and cannot accept as part of a new treaty. (Online EU news)

Ferber is working to achieve a compromise *before the end of* the month. (Online EU news)

The fact that there are more place reference bundles in the EU discourse corpus seems to be in contrast with this kind of explanation. A reason behind it might be the greater variety of places news texts refer to. As illustrated in Tables 11 and 12, the lexical bundles in the place reference category of news texts refer to three countries around the world and one EU institution in the online EU news texts, whereas bundles in the EU texts in the same category refer to member states of the EU, the Official Journal of the EU and places in general:

The text will *be sent to the* European Parliament for a second reading, in the framework of the codecision procedure. (EU discourse)

It reviews the main trends across the EU and *in the Member States*.(EU discourse)

Buried by voters *in France and the* Netherlands in 2005, it has been resurrected by the EU's German presidency and put on the table of a summit in June. (Online EU news)

Categories	Sub-categories	Bundles
I. Stance bundles Attitudinal/ modality	Desire	would like to see
	Obligation/ directive	for the EU to will have to be
	Intention/ Prediction	the EU wants to there will be a
II. Discourse organisers	Topic elaboration/ clarification	on the other hand
	Structure signal	said in a statement
III. Referential bundles Specification of attributes	Quantity specification	# per cent to per cent in the the rest of the
	Tangible framing	the size of the both sides of the
	Intangible framing	on the basis of as a result of in the light of the idea of a the substance of the as part of the
Time/Place/ Text reference	Place reference	in the Council of in France and the France and the Netherlands and the United States
	Time reference	when it comes to at the end of the end of the at the same time by the end of for the first time at a time when before the end of end of this year the end of this at the time of end of the year the future of the
	Multi-functional	the European Union on
IV. Subject-specific bundles EU-related bundles	Person	German Chancellor Angela Merkel president of the European Commission President Jose Manuel President Jose Manuel Barroso Prime Minister Tony Blair European Commission President Jose Prime Minister Romano Prodi British Prime Minister Tony
	Country/ organisation	the EU and the of the European Union the Council of Ministers presidency of the EU EU and the US and the Czech republic between the EU and

		the European Commission is the European Court of the European Union and the European Central Bank of the European Commission that the EU is the European Commission to and the European Commission court of first instance European Court of Justice of the European Parliament of the United States the EU trade commissioner
	Event/ document	the Treaty of Rome charter of fundamental rights 50th anniversary of the the 50th anniversary of for a new treaty of the treaty of the charter of fundamental
	Other	French and Dutch voters by French and Dutch joined the EU in said that the EU justice and home affairs negotiations on a new rejected by French and and Dutch voters in bid to join the the rejection of the
Not EU-related		the World Trade Organization the support of the

Table 11. Lexical bundles in the Online EU news corpus classified according to their discourse functions

3.5 Subject-specific bundles

This category was added to the original taxonomy (Biber, *et al.*, 2004) following Hylandø (2008) revised version in which he used a category of topic where he classified lexical bundles which were related to the field of research. In the case of the EU-related registers, the subject-specific category was introduced to identify bundles which are related to the issues in connection with the European Union, on the one hand, and general world issues, on the other. The subject-specific category was divided into two main sub-categories of EU-related and Not EU-related. The EU-related category was subdivided into the following sub-categories: Persons/positions, Organisations/countries, Events/documents and Other.

In general, as can be seen in Figure 4, there are more subject-specific bundle types in Online EU news and these are distributed in more categories. Lexical bundles in the Online EU news corpus are represented in all sub-categories of subject-specific bundles. An example for the not EU-related bundles is given here:

But since trade negotiations at *the World Trade Organization* collapsed last July, the EU has set out a strategy for concluding individual and regional trade accords. (Online EU news)

The other sub-category where no bundles from EU texts can be found is persons/positions. There are several European country leaders and European Union leaders mentioned in online EU news texts:

German Chancellor Angela Merkel and *European Commission President Jose Manuel Barroso* will get President Bush's support for a proposal to boost trans-Atlantic commerce by eliminating some bureaucratic hurdles. (Online EU news)

British Prime Minister Tony Blair said the leaders complete the toughest part of the process. (Online EU news)

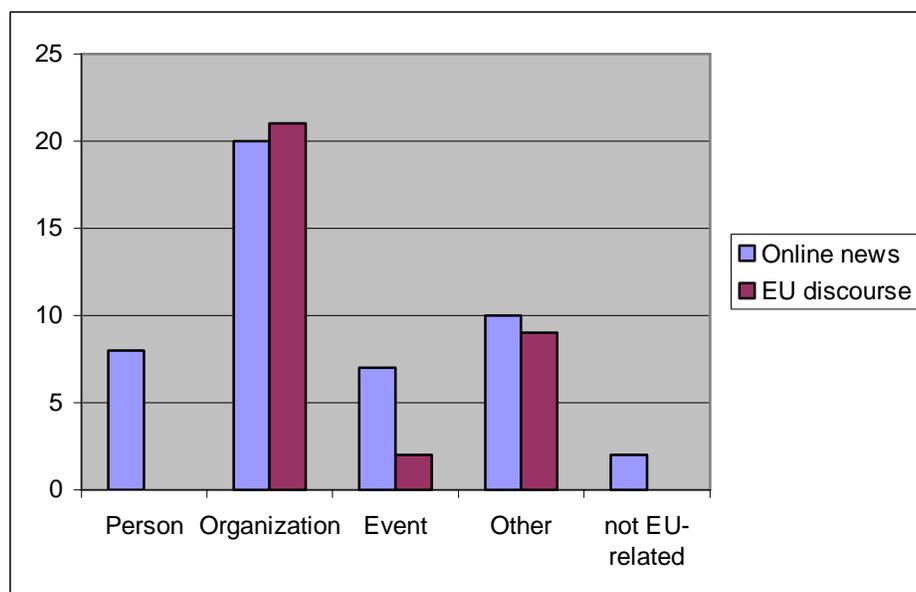


Figure 4. Subject-specific bundles across registers

A less EU-oriented picture develops as bundles in the other subject-specific sub-categories in online EU news texts are analysed and compared to lexical bundles in the same categories in EU texts. Under Organisations/countries institutions and member states of the EU prevail in the EU discourse corpus, whereas in the Online EU news corpus besides the main EU institutions reference to specific countries is made:

Merkel has also offered some support for Bush in a dispute with Russia over U.S. plans to deploy a missile defense system in Poland *and the Czech Republic*. (Online EU news)

The European economy will grow more quickly than that *of the United States* this year, the European Commission claimed yesterday as it revised up its GDP forecasts. (Online EU news)

The governments of the *Member States and the* European Commission were represented as follows: (EU discourse)

The Council confirmed an agreement with *the European Parliament and the* Commission on statements regarding the budgetary management of Community agencies (8636/07). (EU discourse)

Categories	Sub-categories	Bundles
I. Stance bundles Attitudinal / modality	Obligation/ directive	it is important to the member states to
II. Discourse organisers	Topic introduction	it should be noted should be noted that the commission concluded that
	Topic elaboration/ clarification	in accordance with the set out in the as last amended by as set out in as well as the in line with the in addition to the on the protection of last amended by regulation in accordance with article in order to ensure published in the official establishing the European Community of the council on take into account the to ensure that the
III. Referential bundles Specification of attributes	Identification/ focus	main elements of the
	Quantity specification	substantial part of it
	Tangible framing	in the form of
	Intangible framing	on the basis of in the context of in the light of the basis of a the basis of the within the scope of as part of the fall within the scope in view of the with regard to the the application of the the results of the the development of the in relation to the the development of a the opinion of the
	Quality specification	a high level of high level of protection
Time / Place / Text reference	Place reference	be sent to the in the member states in the official journal be obtained from the

		in all member states be published in the in each member state
	Time reference	at the time of at the end of at the same time from the date of
	Text-deixis	provided for in this
	Intertextual reference	having regard to the
	Multi-functional reference	by the commission on the commission on the
IV. Subject-specific bundles EU-related bundles	Country/ organisation	of the European Union official journal of the of the member states minister of state for member states and the of the European Parliament the European Parliament and and of the Council European Parliament and of Parliament and of the the member states and the official journal of Council of the European EC of the European of the Council of the Commission and the the Commission of the the Council of the and the United Kingdom Ministry of Economic Affairs to the member states
	Event/ document	the treaty establishing the treaty establishing the European
	Other	justice and home affairs on health and safety the implementation of the not significantly impede effective significantly impede effective competition would not significantly impede effective competition in the for the implementation of impede effective competition in

Table 12. Lexical bundles in the EU discourse corpus classified according to their discourse functions

Fisher's exact statistical test comparing the number of lexical bundles of different functional types (see Table 13) resulted in a probability value of $p = 0.000$, which shows that the lexical bundles of the two corpora are significantly different regarding the discourse functions they serve in the respective EU-related registers.

Categories	Online EU news	EU discourse
I. Stance bundles	5	2
II. Discourse organisers	2	19
III. Referential bundles	29	36
IV. Subject-specific bundles	47	32
Total:	83	89

Table 13. Number of lexical bundles in the two corpora according to functional categories

3.6 Summary and discussion

The following major distributional patterns emerge from the preceding comprehensive analysis of single-word lexical items and four-word lexical bundles in the two EU-related registers:

- Patterns among the most frequent 50 single-word items in both corpora resemble patterns found in specialised corpora.
- The high frequency of content words and abbreviations in the EU discourse corpus suggests a highly technical language use.
- Lexical bundles are common in both EU-related registers.
- Lexical bundles have a relatively high proportion of total words in official EU texts, which suggests a highly formulaic language use.
- Lexical bundles identified in the EU-related registers are markedly different from lexical bundles of other registers.
- Less than ten percent of lexical bundles identified in official EU texts are used in online EU news texts.
- Many lexical bundles in official EU texts incorporate verb phrases with passive verbs that are characteristic of technical language use.
- Stance is the least frequent function lexical bundles perform in official EU texts.
- Discourse organising is the least frequent function lexical bundles perform in online EU news texts.
- Considerably more bundles serve as discourse organisers in official EU texts.
- The majority of bundles serve referential functions in official EU texts.
- Most lexical bundles are subject-specific in the online EU news texts.

All these patterns describe the EU-related registers in question individually. A comparison of distributional patterns of single-word lexical items and four-word lexical bundles in the two EU-related registers is summarised in Table 14. This comparative analysis suggests that the two EU-related registers differ considerably. The findings support earlier studies (Biber, & Barbieri, 2007; Cortes, 2004; Hyland, 2008) which show that there is a marked contrast in the frequently applied forms, structures and functions across registers, and extend these by examining non-academic registers and related linguistic variation to the third aspect of the Hallidayan (1978) concept of register, that is, tenor, the social relations between participants in communication. Previously comparative analyses were conducted on registers different in mode, i.e. written versus spoken (Biber, & Conrad, 1999; Biber, & Barbieri, 2007) or in field, i.e.

academic writing in different disciplines, such as history and biology by Cortes (2004) and biology, electrical engineering, applied linguistics and business studies by Hyland (2008). The present study, however, investigated two registers that were both written, both discussed EU topics, but differed in tenor, more specifically in their authors and audience. Although the small size of the corpora does not allow for generalisation, the findings seem to suggest that differences in tenor are also important factors in linguistic variation and a frequency-based approach can be an appropriate research framework for this kind of analysis.

	EU discourse	Similarities	Online EU news
Frequency of single-word items	<ul style="list-style-type: none"> twice as many content words twice as many nouns three times more abbreviations 		
Lexical bundles	<ul style="list-style-type: none"> higher proportion of total words in bundles 	<ul style="list-style-type: none"> total number of lexical bundle types 	
	<ul style="list-style-type: none"> more EU perspective 	<ul style="list-style-type: none"> 9 bundles the same 4 EU register specific bundles the same 	<ul style="list-style-type: none"> more global and country-specific perspective
Structure of lexical bundles	<ul style="list-style-type: none"> more VP bundles twice as many VP bundles with non-passive verb ten times more VP bundles with passive verb 	<ul style="list-style-type: none"> most bundles incorporate noun phrase or prepositional phrase fragments 	<ul style="list-style-type: none"> more dependent clause bundles
	<ul style="list-style-type: none"> the distribution of lexical bundles across structural types in the two corpora is significantly different 		
Function of lexical bundles	<ul style="list-style-type: none"> considerably more discourse organising bundles 		<ul style="list-style-type: none"> more stance bundles
	<ul style="list-style-type: none"> more intangible framing bundles 	<ul style="list-style-type: none"> similar number of referential bundle types 	<ul style="list-style-type: none"> more time bundles
	<ul style="list-style-type: none"> text-related bundles (focus, text-deixis, intertextual) only in this corpus 		
		<ul style="list-style-type: none"> similar number of organisation bundle types 	<ul style="list-style-type: none"> more subject-specific bundles person, event, not EU-related bundles only in this corpus
	<ul style="list-style-type: none"> the distribution of lexical bundles across functional categories in the two corpora is significantly different 		

Table 14. Summary of characteristic features of EU-related registers compared

4. Conclusion

In the light of the findings of the present study many issues need to be taken into consideration when designing materials and courses for English language teaching for EU studies. The scope of the present analysis, including its results is obviously limited to the small corpora analysed. Nevertheless, it has revealed important discourse characteristics of EU texts

and online EU news and has also pointed out considerable differences between the two EU-related written registers.

Results indicate that official EU texts make use of a great deal of formulaic language and seem to be lexically more dense than online EU news. This density might cause difficulties for students who may need more practice to be able to read and understand such texts. Further investigation of a larger corpus of EU texts should result in a set of target bundles that can be taught to students explicitly in order to give them this practice.

As for the appropriateness and usefulness of online EU news texts for EU English instruction, the findings suggest that their application is limited to certain aspects of vocabulary teaching. As can be seen in Table 13 and 14, online EU news texts have more subject-specific bundles. Although these lexical items reflect a more global perspective, as these provide a wider range of EU-related vocabulary, it makes them suitable for language instruction.

On the other hand, experts working in an EU context should be familiar with specific vocabulary, EU-specific discourse patterns and text organisation patterns which can only be provided by authentic official EU texts. Further analysis into the discourse of texts used by EU experts in EU institutions on a larger scale could form the basis of course and materials design for English for EU instruction.

Proofread for the use of English by: Robin Bellers, Centre for Academic Writing, Central European University, Budapest.

References:

- AntConc Corpus analysis software. Retrieved August 15, 2007, from <http://www.antlab.sci.waseda.ac.jp/software.html>.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard, & S. Oksefjell (Eds.), *Out of corpora* (pp.181-190). Amsterdam-Atlanta GA: Rodopi.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17(4), 391-406.

- Cowie, A. P. (1992). Multi word lexical units and communicative language teaching. In Arnaud, P. J. L., & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics* (pp.1-12). London: Macmillan.
- Halliday, M. A. K. (1978). *Language as a social semiotic – the social interpretation of language and meaning*. London: Edward Arnold.
- Hunston, S. (2002). *Corpora in applied linguistics*. Oxford: Oxford University Press.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*.27(4), 4-21.
- Lénárt, L. (2003). *EU texts*. Miskolc: Bíbor Kiadó.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies*. London and New York: Routledge.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Moon, R. (2000). Vocabulary connections: Multi-word items in english. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy* (pp.40-63). Cambridge: Cambridge University Press.
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards, & R. Schmidt (Eds.), *Language and communication* (pp.191-230). London: Longman.
- Scott, M., & Tribble, C. (2006). *Textual patterns*. Amsterdam/Philadelphia: John Benjamin Publishing Company.