



OLGA DONTCHEVA-NAVRATILOVA

LEXICAL BUNDLES IN ACADEMIC TEXTS BY NON-NATIVE SPEAKERS

Abstract

Recent research (e.g. Biber et al. 1999, Wray 2000, Biber 2006, 2007, Cortes 2002, 2004, Hyland 2008, Simpson-Vlach and Ellis 2010) has evidenced the importance of lexical bundles – recurrent sequences of words – as a major component in coherent linguistic production and an essential aspect of the shared knowledge of a professional discourse community. While most investigations of lexical bundles in academic discourse have focused on their identification, structure, discourse functions and discipline variation, significantly less attention has been paid to the problems non-native speakers experience in acquiring genre- and discipline-specific recurrent expressions.

The present study is concerned with the use of lexical bundles in non-native speaker academic discourse. The investigation is carried out on a corpus of diploma theses written by Czech students of English with the aim of finding out to what extent the authors of diploma theses use the different functional types of lexical bundles typical of expert academic discourse. While addressing the issues of accuracy, first language interference and the role of different functional types of bundles in coherent discourse production, the author also considers some approaches to teaching genre- and discipline-specific lexical bundles.

Key words

Lexical bundles; discourse functions of lexical bundles; academic discourse; NNS academic writing; diploma theses; L1 interference; accuracy; disciplinary variation

1. Introduction

The increasing amount of corpus-based research into English for academic purposes has evidenced that academic discourse is characterized by a prominent

vocabulary layer and in particular by a large proportion of conventionalized expressions, also called ‘formulaic sequences’ (Wray 2000, 2002), ‘formulaic expressions’ (Simpson 2004), ‘fixed expressions’ (Moon 1992), ‘lexical phrases’ (Nattinger 1988), ‘multiword lexical units’ (Cowie 1992), ‘*n*-grams’ (Cheng et al. 2009), or ‘lexical bundles’ (e.g. Biber et al. 1999, Cortes 2002, Hyland 2008). While the studies into conventionalized expressions apply different explanations and criteria for the identification of multi-word sequences and thus approach their analysis from different perspectives, they all conclusively show that conventionalized expressions constitute a large proportion of discourse and therefore indicate competent use of language in a particular context. This is in conformity with Sinclair’s (1987) ‘idiom principle’ which postulates that speakers and writers do not select the words that they use one at a time, but rather choose units of meaning expressed by pre-constructed phrases. When approaching the study of extended fixed collocations in academic discourse from a pedagogical perspective, it is important to acknowledge that they are a major component of fluent discourse production and comprehension, and represent an essential aspect of the shared knowledge of a professional discourse community. Thus the acquisition of conventionalized expressions used by a target discourse community is an indispensable aspect of the communicative competence that learners must acquire to be fully socialised in an academic setting (cf. Cortes 2004: 398). In addition, as Hyland (2008: 4) points out, “[lexical] bundles are not only central to the creation of academic discourse, but they offer an important means for differentiating written texts by discipline”.

The present investigation into the use of lexical bundles in non-native speaker academic discourse explores the occurrence of highly frequent lexical bundles identified by previous studies (e.g. Biber 2006, Biber and Barbieri 2007, Cortes 2002, Hyland 2008, Simpson-Vlach and Ellis 2010) in a corpus of diploma theses written by Czech students of English. The aim is to find out to what extent the authors of diploma theses use the main functional types of lexical bundles typical of expert academic discourse, while comparing the use of bundles in theses dealing with three fields of study – linguistics, methodology and literature. The results of the quantitative analysis serve as the basis for a discussion of issues of accuracy and first language interference, of the appropriateness of use of different functional types of bundles and their role in coherent discourse production, as well as of some strategies for teaching genre- and discipline-specific lexical bundles.

2. Lexical bundles

The term ‘lexical bundles’ was used for the first time in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999), where it was defined as referring to “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status” (990), and later further specified as the most frequent recurrent sequences of three and more words in a register or genre (Biber and Barbieri

2007: 264). This definition is based on a frequency-driven approach and the frequency cut-off applied in different studies varies, e.g. Biber and Barbieri (2007) use 40 per million words, Hyland (2008) opts for 20 per million words, while Simpson-Vlach and Ellis (2010), who complement frequency with the criterion of cohesiveness, set the cut-off point at 10 per million words.

According to Biber and Barbieri (2007: 269) lexical bundles have several features which distinguish them from other kinds of formulaic expressions: they are extremely common; they are not idiomatic in meaning and lack perceptual salience; and finally, in most cases they do not represent complete structural units, but often bridge structural units, e.g. they begin at a clause/phrase boundary, but the last words of the bundle are the beginning elements of a second structural unit. It seems that this is due to their structural incompleteness and non-idiomatic character that lexical bundles have stayed outside the scope of the lexical expressions traditionally included in foreign language teaching curricula. However, there is now a growing awareness of the necessity of including overt teaching of lexical bundles in EAP curricula; this is evidenced, for instance, by Simpson-Vlach and Ellis's (2010) Academic Formulas List (AFL), an empirically derived list of formulaic sequences for academic speech and writing intended for pedagogical purposes, which comprises a common core of academic formulas shared by multiple disciplines and discipline-specific lexical bundles.

In order to help learners gain control over the use of lexical bundles, it is essential to identify their structural and functional characteristics. Although there are minor differences in the structural types of lexical bundles used in previous research (e.g. Biber 2006, Biber and Barbieri 2007, Hyland 2008), it is possible to assume that there are four major types of bundles:

1. verb phrase components, which can include passive voice, anticipatory *it* structures and dependent clause fragments – *is assumed to be, it is possible that, can be seen as*
2. noun phrase elements followed by a part of a modifier, often an *of*-prepositional phrase – *the end of the, the extent to which, a function of the*
3. prepositional phrases followed by prepositional or clausal elements – *at the end of, of the things that*
4. others, typically longer clausal structures which often function as politeness formulae (Biber 2004: 153) – *as well as the, thank you very much, what are you doing, have a nice day*

Perhaps the most important condition for mastering lexical bundles is an understanding of their discourse functions in a particular register, genre and discipline. As Biber and Barbieri (2007: 270) point out, lexical bundles are “a kind of pragmatic ‘head’ for larger phrases or clauses, where they function as discourse frames for the expression of new information.” The taxonomy of discourse functions of lexical bundles adopted in this research draws on the classifications suggested by Biber, Conrad and Cortes (2004), Cortes (2004), Biber and Barbieri (2007),

Hyland (2008) and Simpson-Vlach and Ellis (2010), and considers three major discourse functions of bundles – referential, discourse organizing and attitudinal – which, as Cortes (2004: 401) points out, reflect ideational, textual and interpersonal meanings (Halliday 1985) respectively. Before moving on to a discussion of the discourse functions of lexical bundles, it should be noted that due to their multifunctionality, which is enhanced by their structural incompleteness, there are some discrepancies between the existing classifications. Apart from differences in the individual super-ordinate categories and their sub-categorizations, there are also differences in the attribution of some realizations to functional categories. Thus, for instance, while Biber (2006) and Simpson-Vlach and Ellis (2010) classify *as a result of* and *on the basis of* as referential expressions specifying the representation of the reality, Cortes (2004), Biber and Barbieri (2007) and Hyland (2008) include them in the group of text organizers establishing logical relations in the discourse. Since some lexical bundles can have multiple functions even within a single occurrence (Biber, Conrad and Cortes 2004: 383, Nesi and Basturkmen 2009: 32), the differences between the above-mentioned categorizations may be attributed to variation in the size and composition of the corpora used by the researchers, which affects the main functional specialization of bundles in the respective registers, genres or disciplines. Thus corpora with a larger hard sciences component seem to yield a higher rate of referential expressions, while the more argument-based character of the humanities and social sciences results in a higher proportion of bundles performing the discourse organizing function. Further evidence of the multifunctionality of lexical bundles is provided, for instance, by the potential of *at the beginning of*, as shown in Figure, to have time, space and text deictic and/or discourse organizing function. However, despite the variation in the pragmatic meanings of lexical bundles, they are typically assumed to fulfil one main function, which in some cases may be discipline- or genre-specific.

The functional classification of lexical bundles adopted in the present research comprises the following three types of bundles, further divided into several sub-categories:

1. Referential bundles (broadly overlapping with referential expressions in Biber and Barbieri (2007) and research oriented bundles in Hyland's (2008) classification) express ideational meanings related to the representation of reality:
 - a) time/place/text-deixis bundles – *at the end of the*, *and the beginning of*
 - b) attribute bundles (specifying procedure, quantity or description of reality) – *a little bit of*, *the use of the*, *the structure of the*
 - c) topic-specific bundles – *in the curricula of*, *the interpretation of the*
2. Discourse organizers (the term is borrowed from Biber and Barbieri (2007); it corresponds to text-oriented bundles in Hyland's (2008) list) convey textual meanings concerned with the organization of a text and the development of argumentation:

- a) logical relations bundles
 - transition bundles (addition/contrast) – *on the other hand, in addition to the, in contrast to the*
 - resultative signals – *as a result of, it was found that, these results suggest that*
 - b) intratextual reference bundles (organizing smaller and larger stretches of discourse) – *in the present study, in the next section, as shown in Figure*
 - c) framing bundles (focusing, limiting conditions on arguments) – *in the case of, with respect to the, in the presence of, with the exception of*
3. Attitudinal bundles (corresponding to Biber and Barbieri's (2007) stance expressions and to Hyland's (2008) participant-oriented features) convey interpersonal meanings including:
- a) stance bundles (express the evaluations and attitudes of the writer) – *the fact that the, it is possible to, are more likely to*
 - b) interactional bundles (address readers and involve them in the argumentation) – *it should be noted that, as can be seen*

It is evident that some structural types of lexical bundles tend to perform specific discourse functions; thus referential time/place/text-deixis and discourse organizing framing bundles tend to be realized by prepositional phrases, discourse organizing resultative and interactional bundles by verbal phrases, and referential attribute bundles by noun phrase elements. It is therefore not surprising that unlike in the highly interactional discourse of conversation, where the majority of lexical bundles are stance expressions containing a verbal element, in academic discourse almost 70 per cent of the most common bundles consist of noun phrase elements or prepositional phrases performing the function of discourse organizers and referential expressions (Biber 2006).

Although it is not possible to claim that lexical bundles contribute directly to cohesion and coherence in discourse, they undoubtedly facilitate discourse processing by organizing discourse in a lesser number of larger meaningful units. Moreover, as pointed out by Nesi and Basturkmen (2009: 26), some of the functions of lexical bundles discussed below may be seen as comparable to conjunctive relations as described in Halliday and Hasan's (1976) model of cohesion. Since coherent discourse interpretation is constructed simultaneously on all planes of discourse (Dontcheva-Navratilova 2009a), all three functional types of lexical bundles can be seen as discourse signals which facilitate accurate understanding and ease fluent discourse production by contributing to the perception of continuity in discourse: attitudinal bundles help build up a consistent authorial voice, referential bundles promote topic continuity, while discourse organizers have a clear cohesive role related to signalling relations between adjacent and distant stretches of discourse.

Before moving to a description of the material under investigation and methodological issues, it is necessary to stress that apart from the common core lexical

bundles which transcend disciplinary variation, each discipline and genre employs a distinctive set of lexical bundles associated with the subject-matter of that discipline and the typical communicative purposes of that genre. Therefore, an investigation into the use of bundles in non-native speakers' writing should take into consideration the acquisition of both common core and discipline-specific lexical bundles.

3. Material and method

In agreement with the view that quantification "should be treated as a starting point of investigation" (Hunston 2007: 46), this study is based on a quantitative analysis of the occurrences of selected lexical bundles in a learner discourse corpus, complemented with a discussion of the accuracy and appropriateness of learners' use of bundles, and possible approaches to teaching common core academic formulas and discipline- and genre-specific bundles. The value of studying formulaic expressions on a frequency basis has been questioned by Wray (2002), who claims that it has not been categorically proved that any word sequences are truly pre-fabricated, i.e. "stored and retrieved whole from memory" (Wray 2002: 9). Nevertheless, since it is generally agreed that lexical bundles are useful devices for the comprehension and construction of discourse which are indispensable for achieving native-like competence and fluency, it seems reasonable to regard a corpus-based frequency analysis as a useful tool for improving our understanding of lexical patterning in different genres and disciplines (cf. Biber and Barbieri 2007: 283).

Within the corpus-based investigations applying a quantitative approach to the study of language, there are two basic methods used for the identification of fixed recurrent expressions. The first approach relies on a search tool that identifies lexical co-occurrences of different length, at different cut-off frequency points (e.g. Biber et al. 1999, Hyland 2008). The second approach, which is adopted in the present study, explores sets of expressions selected prior to the study (e.g. Nattinger and DeCarrico 1992) which may be considered representative of a particular discipline or genre based on the results of previous research and native speakers' perception of appropriateness. As it allows an assessment of the extent to which foreign language learners have acquired register, discipline and genre-specific target bundles, this approach seems to be particularly appropriate for the analysis of non-native speakers' academic discourse.

This investigation into lexical bundles in academic texts by non-native speakers studies 4-word bundles, since their functional specification tends to be clearer than that of 3-word bundles, and they are more frequent and show less variation than 5- and more-word bundles. (It should be noted that 3-word bundles are often included in 4- and more word bundles). In addition, this is the approach taken in most previous studies and thus will allow a comparison of the findings of this investigation with those of previous research. The target bundles were selected

on the basis of the results of previous research into lexical bundles in similar genres and disciplines, i.e. Hyland's (2008) research into bundles used in doctoral and Master's theses in electrical engineering, biology applied linguistics and business studies, Cortes's (2004) and (2006) analysis of published and students' academic writings in the fields of biology and history, and Simpson-Vlach and Ellis's (2010) Academic Formulas List based on an analysis of expert academic discourse.

The present analysis is carried out on a corpus of fifteen Master's degree theses in the fields of linguistics, methodology and literature, written by Czech students of English following teacher-training study programmes at Masaryk University in the period from 2005 to 2008. It should be noted that since these programmes include an academic writing course focused on basic academic genres and their lexicogrammatical features and an elective Academic Skills course with a similar orientation, the authors of the theses are expected to have acquired some academic writing skills. In addition, when writing their Master's theses students take part in tutorials conducted by their thesis supervisors, who tend to highlight discipline-specific genre conventions.

The text of the theses included in the present investigation was converted into an electronic corpus with three sub-corpora – one representing linguistics theses (88,000 words), the second methodology theses (88,000 words) and the last literature and cultural studies theses (78,500 words); thus the overall size of the material under investigation is approx. 254,000 words. The analysis takes into consideration only text written by the authors of the theses, i.e. the text of quotes, citations (including summaries where the stance marker is used by the original author) and examples has been excluded from the material. The corpus was searched for the target lexical bundles using MicroConcord (Scott et al. 1993), a concordance programme. Since the present research analyses non-native speaker discourse typically characterized by a lower frequency of occurrence of academic formulas, the frequency cut-off for considering that a lexical expression has reached bundle rate was set at 20 occurrences per million words. To avoid the common limitations of small corpora, such as the influence of authors' idiosyncrasies and the use of normed rates based on a low number of raw occurrences, in agreement with previous research an additional distributional requirement was introduced according to which target structures should occur in four out of the five diploma theses in at least one of the sub-corpora to be considered as yielding significant frequency rates.

The results of the quantitative analysis are considered both from a structural point of view, in terms of the learners' competence in using the grammatical types of bundles, and from a functional point of view, according to the ability of the students to use bundles as means for enhancing fluency and discourse coherence. The findings are compared to the data reported by the studies used for identifying the target bundles, based on normalized frequencies per million words.

4. Analysis and discussion

This investigation focuses on 30 4-word bundles (target bundles), which were chosen to represent the most frequent bundles used in academic discourse in the field of humanities. (Despite some differences in the results reported in different studies, the frequency of occurrence of these bundles is assumed to exceed 50 per million words.) In addition, since according to Simpson-Vlach and Ellis's (2010) Academic Formulas List the majority of the bundles included in this research are core academic bundles, it was assumed that these were the bundles that non-native writers were most likely to have acquired during their university studies. Topic-specific referential bundles were excluded from the research, since they are subject-specific and thus not representative of the general academic writing skills of Czech writers of diploma theses.

The target bundles are listed in Table 1, grouped according to form and function. Since the selection is frequency-based, it does not represent evenly all the structural and functional types of lexical bundles. The most numerous structural types of bundles are prepositional phrase structures (13 bundles) and noun phrase elements (8 bundles); with regard to their functional classification, referential and discourse organizing bundles are the more numerous groups (13 and 11 bundles respectively), while attitudinal bundles are less common (6 bundles). It should be noted that in the present research multifunctional bundles are classified according to what is considered to be their main function; however, in some cases this may affect the results of the quantitative analysis, since not only can bundles perform different functions in different contexts, but, as mentioned above, they can have more than one function within a single occurrence.

Table 1. Target bundles

Types of bundles	Referential LBs	Discourse organizing LBs	Attitudinal LBs
Verb phrase components		<i>that led to the</i> <i>is due to the</i>	<i>it is important to</i> <i>it is possible that</i> <i>should be noted that</i> <i>can be found in</i>
Noun phrase elements	<i>the extent to which</i> <i>the end of the</i> <i>one of the most</i> <i>the relationship between the</i> <i>the structure of the</i> <i>the role of the</i>	<i>the results of the</i>	<i>the fact that the</i>
Prepositional phrase structures	<i>at the same time</i> <i>at the end of</i> <i>at the beginning of</i> <i>in the form of</i> <i>in the context of</i> <i>in terms of the</i> <i>on the basis of</i>	<i>on the one hand</i> <i>on the other hand</i> <i>and on the other</i> <i>in the present study</i> <i>in the next section</i> <i>in the case of</i>	
Other expressions		<i>as well as a/the</i> <i>as a result of</i>	<i>I would like to</i>

Table 2 below shows the raw frequency of occurrence of the target bundles in the material under investigation, while reflecting the functions of the bundles and the preferences of the authors of the diploma theses according to their fields of study – linguistics, methodology and cultural studies and literature. It is evident that the writers of the diploma theses use a limited repertoire of lexical bundles and that many bundles are rarely used. This tendency is particularly prominent in the cultural studies and literature theses, where eleven out of the 30 bundles do not occur at all, and in the methodology theses, where seven of the target bundles are never used. Linguistics theses show the widest repertoire and the highest rate of lexical bundles, which may be tentatively explained by a higher degree of language awareness on the part of students with an interest in linguistics. It should be noted that previous research carried out on the same corpus has evidenced that authors of linguistics theses also show a strong tendency towards explicit marking of stance and logical relations, which in the case of some stance adverbials (Dontcheva-Navratilova 2008, 2009b) and contrastive discourse markers (Povolná 2010a, 2010b) even exceeds the standard rate in native-speaker discourse. According to Vogel (2008), who explores a slightly different corpus of Czech students' essays in English, the higher rate of adverbial sentence linkers in texts by non-native speakers may result from exposure to overt teaching of the target structures and from the comfort stemming from the use of explicit discourse markers. In addition, the frequent use of a specific bundle or of a certain type of bundle may be also seen as a form of overgeneralization (Cortes 2004: 412), i.e. as a stage in the process of acquisition of academic discourse, during which novice writers use a high rate of bundles before learning how to use them in an appropriate and balanced way.

Table 2. Raw frequencies of occurrence of the target bundles in the corpus

Target bundles	Linguistics	Methodology	Literature	Total No
Referential LBs				
<i>at the same time</i>	13	20	4	36
<i>in terms of the</i>	4	3	2	9
<i>on the basis of</i>	7	4	0	11
<i>in the form of</i>	24	3	3	30
<i>in the context of</i>	1	2	2	5
<i>the extent to which</i>	3	0	0	3
<i>one of the most</i>	10	7	4	21
<i>the end of the</i>	7	4	17	28
<i>at the end of</i>	14	7	18	39
<i>at the beginning of</i>	9	15	7	31
<i>the relationship between the</i>	3	5	0	8
<i>the structure of the</i>	2	1	2	5
<i>the role of the</i>	2	7	11	20
Referential total No	99	78	70	247
Discourse organizing LBs				
<i>on the one hand</i>	4	1	4	9
<i>on the other hand</i>	42	23	23	88
<i>and on the other</i>	1	1	0	2
<i>in the present study</i>	2	0	0	2
<i>in the next section</i>	0	0	0	0

Target bundles	Linguistics	Methodology	Literature	Total No
<i>in the case of</i>	15	0	1	16
<i>the results of the</i>	4	10	0	14
<i>that led to the</i>	1	0	0	1
<i>is due to the</i>	2	0	0	2
<i>as well as the</i>	7	2	5	14
<i>as a result of</i>	2	2	4	8
Disc. organizers total No	80	39	37	156
Attitudinal LBs				
<i>it is important to</i>	9	15	2	26
<i>it is possible that</i>	0	5	0	5
<i>should be noted that</i>	4	0	0	4
<i>can be found in</i>	7	1	2	10
<i>the fact that the</i>	17	4	6	27
<i>I would like to</i>	23	5	4	32
Attitudinal total No	60	30	14	104
All types total No	239	147	121	507
All types total %	47	29	24	100

An analysis of structural variation in the use of lexical bundles by the authors of the diploma theses indicates that prepositional phrases and noun phrase elements show the highest rate of occurrence. While the occurrence of 4-word bundles shows a lower frequency rate, 3-word bundles, which are part of larger units, are extensively used, e.g. *the fact that* (128) vs. *the fact that the* (27); *in terms of* (47) vs. *in terms of the* (9); *of the most* (46) vs. *one of the most* (21); *the form of* (45) vs. *in the form of* (30). At the same time, some 4-word sequences form parts of larger, 5- and 6-word bundles (although there is some variation in the collocates, which can often be regarded as synonymous or related to a superordinate concept), which tend to show relevant frequencies, especially in the methodology and literature and cultural studies sub-corpora, and thus may be considered as discipline-specific. For instance, in literature and cultural studies theses the bundles *(at) the end of the* and *at the beginning of* tend to collocate with lexical items referring to text types (e.g. *novel, story*) and time periods (e.g. *century, year, (past) decade*), while in methodology theses the same bundles tend to co-occur with words referring to teaching units (e.g. *course, lesson, class*). Another example is provided by the bundle *the role of the*, which in the methodology corpus collocates with *teacher*, and in the literature and cultural studies corpus with lexical items referring to characters (e.g. *character, mother, children*). In linguistics theses, there is variation in the structure of the attitudinal bundles *it should be noted/stated/remarked that* and *I would like to emphasise/stress/mention/stress/summarize*, which express the evaluative stance of the writer.

As to the issue of accuracy, the results of the analysis show that there are some occasional grammatical errors in the use of lexical bundles, which raises questions about the students' awareness of the bundle status of a given fixed expression. This concerns mainly the use of the definite article in the bundles *in the terms of*, *on the one hand* and *in the case*, where in the first bundle the definite article is inserted unnecessarily, while in the second and the third it is frequently

omitted. Owing to the lack of articles in the Czech language system, accuracy problems concerning the use of both the definite and indefinite articles are common in Czech speakers of English; however, it is precisely an awareness of the fixed structure of bundles that can help Czech learners improve their English language fluency in this regard.

Apart from structural variation and grammatical accuracy, an important criterion for the assessing of the extent of bundles acquisition is the functional appropriateness of the use of the target expressions by the Czech writers of diploma theses. Tables 3, 4 and 5 below show the normalized frequencies of occurrence per million words of the functional types of target bundles to allow a comparison between the rate of bundles in native-speaker writing and academic texts by non-native speakers. Since in this investigation the criteria for acknowledging bundle status include a frequency rate cut-off of 20 per million words and occurrence in at least four out of the five diploma theses included in each of the three sub-corpora, it is evident that very few of the target expressions have significant frequency rates in the material; these are highlighted in bold in Tables 3, 4 and 5. (The instances of relatively high normalized rate of expressions which are not regarded as bundles is due to overuse of these target expressions by some authors. The preference of students for a recurrent use of expressions that they know and can use correctly is also evidenced by Povolná’s (2012, in this issue) investigation into the use of discourse markers in the same corpus.) It follows that in general writers of diploma theses tend to use bundles considerably less frequently than native-speaker writers.

The first functional type of lexical bundles, comprising two major sub-categories (time/place/text deixis bundles and attribute bundles), is represented by referential expressions which typically contribute to the description of research objects and contexts, research procedures and methodology. While all deictic expressions used in the corpus of diploma theses can be regarded as showing significant frequencies at least in some of the sub-corpora and one in the whole corpus, none of the attribute phrases used for physical description (*the structure of, in the form of, in the context of*), quantification (*the extent to which, one of the most*) and outlining a procedure (*the role of the, on the basis of, in terms of the, the relationship between*) has the frequency of occurrence required for the acquiring of bundle status in more than one of the corpora (Table 3).

Table 3. Referential bundles – normalized frequencies per million words

Target referential bundles	Linguistics	Methodology	Literature	Average frequency
Time/place/text deixis LBs				
<i>at the same time</i>	148	227	51	145
<i>the end of the</i>	80	45	217	110
<i>at the end of</i>	159	80	229	153
<i>at the beginning of</i>	102	170	89	122

Target referential bundles	Linguistics	Methodology	Literature	Average frequency
Attribute LBs				
<i>in terms of the</i>	45	34	25	35
<i>on the basis of</i>	80	45	0	43
<i>in the form of</i>	273	34	38	118
<i>in the context of</i>	11	23	25	20
<i>the extent to which</i>	34	0	0	12
<i>one of the most</i>	114	80	51	83
<i>the relationship between the</i>	34	57	0	31
<i>the structure of the</i>	23	11	25	20
<i>the role of the</i>	23	80	140	79

As already mentioned, the bundles (*at the end of the* and *at the beginning of*) can be considered discipline-specific in literature and cultural studies diploma theses, where they typically have a time-deictic (1) and text-deictic function (2). However, there is one occurrence of *at the end of* with text-deictic and discourse organizing function in the linguistics sub-corpus (3).

- (1) Dorsey's involvement in gospel chorus music was caused by two factors: firstly, choirs played a significant role in the conservative black Baptist and Methodist churches, and secondly, his lack of success with promoting his gospel solo songs at the end of the 1920s and the beginning of the 1930s. (Literature)
- (2) At the end of the novel the situation is solved by Harriet's accidental death. (Literature)
- (3) At the end of the theoretical part, I will discuss methodology issues connected with the questionnaire used in my research. (Linguistics)

The multifunctionality of lexical bundles is also illustrated by the expression *at the same time*, which apart from its temporal meaning indicating simultaneity (4) can be used metaphorically to indicate addition (5) or concession (Quirk et al. 1985). In harmony with the findings of Cortes (2004: 413), in the material *at the same time* is used to indicate both simultaneity and addition. However, there are no instances of this bundle with concessive meaning, which confirms that concessive and contrastive relations are "the most complex of all semantic relations that may hold between parts of a discourse" (Kortmann 1991: 161).

- (4) In other words, the procedure consists of hearing and seeing a word at the same time, then saying it, then it should be covered and a learner writes the word down and checks spelling. (Methodology)
- (5) Spoken legal English is not just a spoken variant of the written text. It is a different genre at the same time because there is a very tight connection between what is said, how it is said and why, and the situation in which the speech is uttered. (Linguistics)

Only three attribute referential bundles show significant frequency rates. The most prominent expression is *in the form of*, which can be seen as discipline-specific, as it is used only in the linguistics theses to refer to the realization of structural and functional categories, as in:

- (6) Nevertheless, *I mean* seems to share enough features with *you know* and *you see* (e.g. they all take a direct object in the form of a nominal that-clause, they all belong to the “private” type of factual verbs and they help the smooth flow of conversation) that it is possible to include it into the category of comment clauses. (Linguistics)

The phrase *on the basis of* shows relevant frequency, but it is used only in two linguistics and two methodology theses to indicate a classificatory criterion (7). It is interesting to note that the shorter 3-word bundle *on the basis* shows exactly the same extent of occurrence, i.e. the structure *on (the) basis (that) + finite clause*, which is regarded as an adverbial subordinator in the process of ongoing grammaticalization (Mair 2010), does not occur in the corpus under investigation. This may be explained by a lack of overt instruction in structures which are still undergoing grammaticalization and by the preference of non-native novice writers for the use of well-established and functionally unambiguous expressions.

- (7) Inspired by Tottie, I decide to categorize long backchannels on the basis of the most significant and the most frequent backchannel item of the particular backchannel. (Linguistics)

The bundle *one of the most* is used with significant frequency in the linguistics theses to express quantification and to highlight a specific aspect of the research object, as in (8):

- (8) One of the most visible features of spoken informal language in the advertisements is the use of contractions. (Linguistics)

As already mentioned above, the last statistically prominent expression *the role of the* may be considered as specific to the methodology sub-corpus, where it indicates the functional categorization of participants (the teacher in particular) in classroom interaction (9), and to the literature sub-corpus, where it indicates the roles assumed by characters in the fictional world (10). However, while occurrences of the 3-word bundle *the role of* are genuinely numerous in all literary and methodology theses included in the corpus, the 4-word bundle *the role of the* was actually found only in four literature and two methodology theses.

- (9) The role of the teacher during these activities can be: a group process manager, a facilitator, a counselor (answer questions and monitor the activity), an independent participant within the group. (Methodology)

- (10) Sue with Tom start to fulfill the role of the children and the mood in the house changes as well. (Literature)

The second functional type of lexical bundles – discourse organizers – is used in expert writing for connecting ideas, building up coherent argumentation and guiding readers through a text. As the findings of the analysis show, the three sub-categories of discourse organizers are not represented in a similar way in the diploma theses by Czech students of English (Table 4). Contrary to Hyland's (2008) data based on doctoral and Master's theses in English written by Cantonese L1 speakers, the rate of intratextual reference bundles in the material under investigation is very low and is actually restricted in the linguistics theses to two occurrences of *in the present study*. The underuse of intratextual reference bundles may indicate a lack of ability on the part of Czech novice writers to use structuring signals to organize discourse and maybe even a lack of awareness of the necessity of controlling the continuity of subject-matter, of anticipating the expectations of their readers and of helping them follow the argumentation across larger stretches of text. This may be explained by interference from L1 academic writing conventions, since despite the changes which Czech conventions of academic writing are currently undergoing as a result of the globalization of academic discourse, there are cross-cultural differences between the Anglo-American and Central European traditions and norms in academic writing, which, apart from differences in cultural and historical backgrounds, reflect the different grammatical structures of Czech and English (Chamonikolasová 2005, Stašková 2004). Thus, while Anglo-American academic discourse, which is oriented towards the reader, is characterized by careful organization and extensive use of rhetorical and cohesive devices guiding the reader through the argumentation, Czech academic texts, which are orientated towards the topic, tend to be structured less transparently, with the result that devices for guiding the reader through the discourse are rather scarce (Čmejrková et al. 1999: 25-30, Chamonikolasová 2004: 83). The absence of occurrences of *in the next section* may be also explained by the tendency of Czech writers to refer to the structural parts of their theses by the labels 'chapter' and 'part' instead of 'section' and 'subsection'.

Table 4. Discourse organizing bundles – normalized frequencies per million words

Target discourse organisers	Linguistics	Methodology	Literature	Average frequency
Logical relations LBs				
<i>on the one hand</i>	45	11	51	35
<i>on the other hand</i>	477	261	293	346
<i>and on the other</i>	11	11	0	8
<i>the results of the</i>	45	114	0	55
<i>that led to the</i>	11	0	0	4
<i>is due to the</i>	23	0	0	8
<i>as a result of</i>	23	23	51	31
<i>as well as the</i>	80	23	64	55
Framing bundles				
<i>in the case of</i>	170	0	13	63

Target discourse organisers	Linguistics	Methodology	Literature	Average frequency
Intratextual reference LBs				
<i>in the present study</i>	23	0	0	8
<i>in the next section</i>	0	0	0	0

The single framing bundle included in this investigation – *in the case of* – is also used mainly in the linguistics theses to refer to specific cases highlighting particular aspects of the argumentation.

- (11) What can perhaps be more important in the case of cars is the demographic segmentation, which is a study of the numerical characteristics of the population. (Linguistics)

The largest group of discourse organizers considered in the present study comprises logical relations bundles, which can be used to mark contrast (*on the one hand, on the other hand, on the other*), addition (*as well as*) and cause-result (*that led to the, is due to the, as a result of, the results of the*) relations. When considered against the overuse of discourse markers by Czech students reported in previous investigations (Povolná 2010a, 2010b, Vogel 2008), the rate of discourse organizing bundles seems rather low. This may be attributed to the preference of students for the use of linking devices explicitly taught in academic writing courses.

The analysis has shown that only two bundles expressing logical relations show significant frequency. The highest rate of occurrence is shown by the bundle *on the other hand*, which is grammaticalized as a contrastive adverbial linker and often occurs in a correlative pair with *on the one hand*. In the corpus under investigation, however, there are very few examples of correlative use, and it is interesting to note that in half of them the article in the first member of the correlative pair is missing, as in:

- (12) On one hand, it may connote pride or individuality and on the other hand, it may connote allergy, filth or fight. (Linguistics)

In most cases, however, *on the other hand* is used on its own in sentence initial position (13) relating the sentence it introduces to the preceding discourse, typically the preceding sentence. In some cases, it occurs in medial position (14), to contrast qualities attributed to entities referred to by the subjects of two consecutive sentences. In two cases, where the contrast relation is established between the clauses of a compound sentence intrasententially, *on the other hand* is enhanced by a second marker of contrast (*but*), as illustrated in (15).

- (13) The children identify themselves with the norm and pursue to fulfil the norm. Every unsuccessful attempt decreases their self-respect. On the other hand, the emotional acceptance and support of the child by the parents and

teachers increases his/her beliefs in the future success as well as it supports its self-acceptance. (Methodology)

- (14) Each hero also has a villain, a foe, who tries to defeat him. The anti-hero, on the other hand, does not necessary have to be a villain, but has some of his features. (Literature)
- (15) The image may be less explicit than the verbal text, but on the other hand, an image has advantage of being able to communicating more things at the same time (Vestergaard & Schröder 1992: 42). (Linguistics)

The 4-word bundle *as well as the* occurs in all sub-corpora, but it does not meet the distributional criterion of being used in four out of the five diploma theses; however, the 3-word bundle *as well as* is used 97 times in the material. This supports the view that Czech novice writers show a marked preference for the use of well-established linking devices the acquisition of which is facilitated by overt instruction.

Discourse organizers expressing cause-result relations generally show a low frequency rate. Since causal and resultative expressions have a key rhetorical function in academic writing – to indicate the argumentative process and the conclusions that the author draws from the study – their lower rate seems to indicate that Czech novice writers have not yet developed sufficiently their rhetorical skills. There is only one resultative bundle that occurs with significant frequency in the methodology sub-corpus. This is the expression *the results of the*, which is used to introduce the writer's interpretation of the outcomes of his/her research (16). Although the resultative expression *as a result of* (17) occurs in all sub-corpora, its frequency is considerably lower. In this case again, the frequency of the 3-word adverbial linker *as a result* is considerably higher, i.e. 21 occurrences in the whole corpus, as compared to 8 occurrences of *as a result of*.

- (16) The results of the test written with the children showed that Martin was able to recognize the written form of the words and could match them with the Czech equivalents. (Methodology)
- (17) This complex book is filled with many hidden messages and intertextual references which all relate to the framing theme – the struggle of the Native Americans to keep their identity, culture, and beliefs as a result of steady Europeanization of North America. (Literature)

The last type of target bundles – attitudinal expressions – convey interpersonal meanings by indicating evaluative judgements and the degree of commitment on the part of the writers to their claims, and by engaging in a dialogue with the reader. The results of the quantitative analysis show that none of the target bundles shows significant frequency of occurrence in all three sub-corpora (Table 5). It is again the linguistics theses that have the highest rate of attitudinal bundles, as

three of the six target expressions (two stance bundles and one interactional bundle) show significant frequency of occurrence; however, the interactional bundle *it is important to* is the most frequent in the methodology theses.

Table 5. Attitudinal bundles – normalized frequencies per million words

Target attitudinal bundles	Linguistics	Methodology	Literature	Average frequency
Stance LBs				
<i>it is possible that</i>	0	57	0	20
<i>(it) can be found in</i>	80	11	25	39
<i>the fact that the</i>	193	45	76	106
<i>I would like to</i>	261	57	51	126
Interactional LBs				
<i>(it) should be noted that</i>	45	0	0	16
<i>it is important to</i>	102	170	25	102

Stance bundles, which are typically used as hedges to indicate uncertainty and caution (Hyland 2008), seem to be underused in the diploma theses of Czech students. There are two stance bundles which show significant frequency in the sub-corpus of linguistics theses; these are *the fact that the* and *I would like to*. Previous functional taxonomies of lexical bundles differ in their categorization of *the fact that the*; Simpson-Vlach and Ellis (2010) classify it as a referential bundle, Cortes (2004) as a framing text organizer, and Biber and Barbieri (2006) as an epistemic stance bundle. While acknowledging that *the fact that the* is essentially multifunctional, in the present research this bundle is regarded as expressing primarily epistemic stance, as it explicitly assesses the information provided as fact and not opinion. Its use in linguistics theses is frequently associated with explaining the reasons for the phenomena observed; often the interpretation of statistically established linguistic facts is hedged to express a lower degree of commitment to claims (18).

- (18) The reason might be the fact that the length of speakers' turns is relatively short in telephone conversations. (Linguistics)

While in academic discourse stance is frequently expressed impersonally, in the material the highest frequency stance bundle is *I would like to*, which in the linguistics theses shows a marked tendency to collocate with discourse verbs, e.g. *emphasize*, *state* and *mention* (19), thus overtly presenting the information conveyed as the personal opinion of the writer and performing an additional discourse-organizing function. The tendency to use the first person singular pronoun for self-reference is strongest in the linguistics theses (366 occurrences), where apart from indicating the purposes and outcomes of the research *I* is also used to describe procedure, to guide the reader through the argumentation and to state claims. The personal bundle *I would like to* in (19) occurs together with the impersonal expression *it is possible to* in the adjacent clause of the same sentence, which further hedges authorial commitment. It should be noted that

although *it is possible to* was not included in this investigation as a target bundle, it occurs 31 times in the linguistics theses, i.e. its frequency rate is 352 per million words.

- (19) In conclusion to the lexical level analysis, I would like to mention that in some studies on legal language it is possible to learn that legal texts are empty of comparatives and superlatives. (Linguistics)

The low rate of occurrence of *I would like to* in the methodology and literature theses may be explained by the fact that their authors use a restricted range of functions of the first person singular pronoun, i.e. primarily for stating the purposes of the theses and explaining methods and procedures. When expressing opinions and discussing their findings, the authors of literature theses show a preference for the use of the inclusive *we*, presenting themselves as readers of the literary text and thus inviting others to share their interpretation of the discourse, while the authors of methodology theses use the first person plural pronoun both for generalizations and for stating their claims. This can be due to L1 interference, as in common with other Slavic languages, in academic writing in Czech the author's identity, even in the case of single-authored texts, tends to be expressed by first person plural pronouns and verb forms (Chamonikolasová 2004: 82).

The very low number of interactional bundles in the material reflects the lower level of dialogicity in novice academic writing; however, it should be noted that stance devices also have a dialogic character. There is only one interactional bundle which shows significant frequency in the linguistics and methodology theses. The bundle *it is important to* is the most prominent in the methodology theses, where it is used to express the writer's judgment on the importance of the point made in their argumentation (20).

- (20) To enhance cooperative way of learning, it is important to enhance the children's friendly relationships. (Methodology)

To conclude the discussion of the use of lexical bundles in the diploma theses of Czech students of English, it remains to be noted that there are clear differences in the distribution of functional types of bundles in the linguistics, literature and cultural studies, and methodology theses (Table 6).

Table 6. Distribution of functional types of lexical bundles in the three disciplines

Functional types	Linguistics		Methodology		Literature	
	Raw No	Per cent	Raw No	Per cent	Raw No	Per cent
Referential LBs	99	41	78	53	70	58
Disc. organizing LBs	80	34	39	27	37	20
Attitudinal LBs	60	25	30	20	14	12
Total	239	100	147	100	121	100

The findings of previous research suggest that while the most frequent functional type of bundle in academic prose is the referential bundle (cf. Biber and Barbieri's (2007) research carried out on a corpus comprising academic texts from both soft and hard sciences), in social sciences the most prominent function of the lexical bundle is that of discourse organizer (Hyland 2008). The reason for this is considered to be "the more discursive and evaluative patterns of argument in the soft knowledge fields, where persuasion is more explicitly interpretative and less empiricist" (Hyland 2008: 16). The results of the present investigation, however, clearly show that the most prominent functional type of bundle in all three sub-corpora is the referential bundle, which indicates that novice writers have not yet acquired discipline-specific discourse conventions. The particularly high percentage of referential bundles in the methodology and literary theses reflects their more descriptive character. The slightly higher percentage of discourse organizing and attitudinal bundles in linguistics theses may be due to an enhanced language and pragmatic awareness on the part of authors who study language from a linguistics point of view. Nevertheless, the rate of discourse organizers in the linguistics sub-corpora falls considerably below the standards of expert academic discourse.

5. Pedagogical implications

As this investigation into the use of lexical bundles in Master's theses by non-native speakers has evidenced, the acquisition of lexical bundles is a long process in which learners have to master both the structure and the functional specialization of lexical bundles typical of expert academic discourse. An important pedagogical implication of this and similar studies is that exposure of university students to a high frequency of occurrence of lexical bundles through academic reading, i.e. unconscious learning, does not result in their acquisition of these bundles (Cortes 2004: 417). One of the reasons for this may be that lexical bundles are very frequent but not perceptually salient. According to Ellis (2002) frequency is a key factor in language acquisition and therefore one would expect language learners to acquire lexical bundles successfully. However, due to the lack of perceptual salience learners may experience some problems in noticing lexical bundles. This is why it may be helpful to deliberately focus the attention of learners on lexical bundles (cf. Cortes 2004, Biber and Barbieri 2007).

Overt instruction in lexical bundles should account for variation in academic discourse and address the identification of discipline- and genre-specific bundles. Apart from the use of corpus-informed lists, such as Simpson-Vlach and Ellis's (2010) Academic Formulas List, in the preparation of teaching materials, overt instruction should incorporate a focus both on structure and discussion of the functions of lexical bundles. The following tasks and activities can be used in courses of English for academic purposes for assistance in the recognition, practice and contextualization of lexical bundles: recognition of LBs in academic discourse

based on frequency of occurrence in texts; pattern practice to develop confidence, including some contextualization; controlled practice, using substitution drills; checking against corpora; pattern analysis and creative use in written performance.

In conclusion, it should be stressed that the emphasis on ready-made conventionalized expressions in academic language courses is motivated by the fact that they have the advantage of both facilitating more efficient retrieval and enabling learners to concentrate on discourse structure rather than on selection of individual words in the course of discourse processing (Nattinger 1988). As Wray (2000) points out, formulaicity benefits both the speaker/writer and the listener/reader by facilitating discourse processing and thus enhancing the perception of discourse coherence.

6. Conclusion

The findings of the present research into lexical bundles in a corpus of Master's theses written by Czech students of English have evidenced that the frequency of use of lexical bundles is lower than that typical of expert academic discourse and that novice writers in a non-native language use a limited repertoire of lexical bundles. While structural inaccuracy in the use of bundles is not very frequent, the distribution of functional categories in the Master's theses corpus differs considerably from the conventions of expert academic writing. This seems to be due to an insufficient level of development of the rhetorical skills of the writers and to interference from L1 writing conventions. Differences in the linguistics, methodology and literature theses in terms of rates of use of target bundles point to variation in what the writers are trying to achieve in their texts and maybe to different degrees of language and pragmatic awareness on the part of the authors. Further research into the discourse of novice writers in a non-native language may focus on the appropriateness of use of the functional types of lexical bundles, paying particular attention to multifunctional expressions and to the contribution of lexical bundles to the perception of cohesion and coherence in discourse.

The pedagogical applications of this research concern the place of lexical bundles in the teaching of academic English. Since the present investigation has – together with several earlier studies – convincingly proved that exposure to expert academic writing does not lead to successful acquisition of lexical bundles typical of academic discourse, there is an indispensable need to find effective ways for including overt instruction in discipline- and genre-specific bundles in courses of English for academic purposes.

Note

This article is part of the grant project 405/08/0866 Coherence and Cohesion in English Discourse, which is supported by the Czech Science Foundation.

References

- Biber, Douglas (2006) *University Language: A Corpus Study of Spoken and Written Registers*. Amsterdam and Philadelphia: John Benjamins.
- Biber, Douglas and Federica Barbieri (2007) 'Lexical bundles in university spoken and written registers.' *English for Specific Purposes* 26, 263–286.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Biber, Douglas, Susan Conrad, and Viviana Cortes (2004) "'If you look at...': Lexical bundles in university teaching and textbooks.' *Applied Linguistics* 25, 371–405.
- Chamonikolasová, Jana (2005) 'Comparing the structures of academic texts written in English and Czech.' In: *Slovak Studies in English* 1. Bratislava: Univerzita Komenského, 77–84.
- Cheng, Winnie, Chris Greaves, Jihn Sinclair and Martin Warren (2009) 'Uncovering the extent of the Phraseological tendency: Towards a systematic analysis of concgrams.' *Applied Linguistics* 30, 236–252.
- Čmejrková, Světa, František Daneš and Jindra Světa (1999) *Jak napsat odborný text*. Praha: Leda.
- Cortes, Viviana (2002) 'Lexical bundles in Freshman composition.' In: Reppen, Randi, Susan M. Fitzmaurice and Douglas Biber (eds.) *Using corpora to explore linguistic variation*. Amsterdam and Philadelphia: John Benjamins, 131–146.
- Cortes, Viviana (2004) 'Lexical bundles in published and student disciplinary writing: Examples from history and biology.' *English for Specific Purposes* 23, 397–423.
- Cortes, Viviana (2006) 'Teaching lexical bundles in the disciplines: An example from a writing intensive history class.' *Linguistics and Education* 17, 391–406.
- Cowie, Anthony C. (1992) 'Multiword lexical units and communicative language teaching.' In: Arnaud, Pierre J. L. and Henri Bejoint (eds.) *Vocabulary and Applied Linguistics*. Palgrave: Macmillan, 1–12.
- Dontcheva-Navratilova, Olga (2008) 'Interaction in Academic discourse: communicative strategies for the expression of stance.' In: Tomášková, Renáta, Sirma Wilamová and Christopher Hopkinson (eds.) *Ends and Means in Language: Communication and Textual Strategies in Mass Media, Commercial and Academic Discourse. Proceedings from a Conference*. Ostrava: Univerzita Ostraviensis, 14–18.
- Dontcheva-Navratilova, Olga (2009a) 'Building up discourse coherence: Creating identities in political speeches.' In: Dontcheva-Navratilova, Olga and Renata Povolná (eds.) *Coherence and Cohesion in Spoken and Written Discourse*. Newcastle: Cambridge Scholars Publishing, 97–119.
- Dontcheva-Navratilova, Olga (2009b) 'Evaluation in non-native writer's academic discourse: Stance devices.' In: Hanušová, Světlana et al. (eds.) *Research in English Language Teacher Education*. Brno: Masaryk University, 33–42.
- Ellis, Nick (2002) 'Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition.' *Studies in Second Language Acquisition* 24, 143–188.
- Halliday, Michael A. K. (1985) *An Introduction to Functional Grammar*. London: Edward Arnold.
- Halliday, Michael A. K. and Ruqaiya Hasan (1976) *Cohesion in English*. New York: Longman.
- Hunston, Susan (2007) 'Using Corpus to Investigate Stance Quantitatively and qualitatively.' In: Englebretson, Robert (ed.) *Stancetaking in Discourse*, Amsterdam and Philadelphia: John Benjamins, 27–48.
- Hyland, Ken (2008) 'As can be seen: Lexical bundles and disciplinary variation.' *English for Specific Purposes* 27, 4–21.
- Kortmann, Bernd (1991) *Free Adjuncts and Absolutes in English. Problems of Control and Interpretation*. London: Routledge.
- Mair, Christian (2010) 'Grammaticalisation of new patterns of clausal subordination: 'on (the) basis (that) + finite clause' and '(on) account (of) + finite clause' in present-day English.' In: Malá, Markéta and Pavlína Šaldová (eds.) ... *For Thy Speech Bewrayeth Thee. A Festschrift for Libuše Dušková*. Praha: Univerzita Karlova v Praze, 153–168.

- Moon, Rosamund (1992) 'Textual aspects of fixed expressions in learners' dictionaries.' In: Arnaud, Pierre J. L. and Henri Bejoint (eds.) *Vocabulary and Applied Linguistics*. Palgrave: Macmillan, 13–27.
- Nattinger, James (1988) 'Some current trends in vocabulary teaching.' In: *Vocabulary and Language Teaching*. London and New York: Longman, 62–82.
- Nattinger, James and Jeanette DeCarrico (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesi, Hilary and Helen Basturkmen (2009) 'Lexical bundles and discourse signalling in academic lectures.' In: Flowerdew, John and Michaela Mahlberg (eds.) *Lexical Cohesion and Corpus Linguistics*. Amsterdam and Philadelphia: John Benjamins.
- Povolná, Renata (2010a) 'Can non-native speakers of English use contrastive discourse markers correctly when writing academic texts.' In: Malá, Markéta and Pavlína Šaldová (eds.) ... *For Thy Speech Bewrayeth Thee. A Festschrift for Libuše Dušková*. Praha: Univerzita Karlova v Praze, 209–231.
- Povolná, Renata (2010b) 'Exploring sequential relations in learner discourse.' In: Ciuk, Andrzej and Katarzyna Molek-Kozakowska (eds.) *Exploring Space: Spatial Notions in Cultural, Literary and Linguistic Studies. Vol. 2. Space in Language Studies*. Newcastle upon Tyne: Cambridge Scholars Publishing, 177–188.
- Povolná, Renata (2012) 'Causal and Contrastive Relations in Novice Academic Writing'. *Brno Studies in English* 38 (2), page numbers – in the present issue.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Scott, Mike, Tim Johns and Simon Murison-Bowie (1993) *MicroConcord*. Oxford: Oxford University Press.
- Simpson, Rita (2004) 'Stylistic features of academic speech: The role of formulaic expressions' In: Connor, Ulla and Thomas A. Upton (eds.) *Discourse in the Professions. Perspectives from Corpus Linguistics*. Amsterdam and Philadelphia: John Benjamins, 37–64.
- Simpson-Vlach, Rita and Nick C. Ellis (2010) 'An academic formulas list: New methods in phraseology research.' *Applied Linguistics* 31 (4), 463–512.
- Sinclair, John (1987) 'Collocation: A progress report.' In: Steele, Ross and Terry Threadgold (eds.) *Language Topics: Essays in Honour of Michael Halliday*. Amsterdam and Philadelphia: John Benjamins.
- Stašková, Jaroslava (2004) 'Options of identity: Authorial presence in research articles abstracts.' In: *Slovak Studies in English 1*. Bratislava: Univerzita Komenského, 201–207.
- Vogel, Radek (2008) 'Sentence linkers in essays and papers by native vs. non-native writers.' *Discourse and Interaction* 1 (2), Brno: Masaryk University, 119–126.
- Wray, Alison (2000) 'Formulaic sequences in second language teaching: Principle and practice.' *Applied Linguistics* 21 (4), 487–489.
- Wray, Alison (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

OLGA DONTCHEVA-NAVRATILOVA is Assistant Professor of English Linguistics at Masaryk University Brno, Czech Republic. She specialises in discourse analysis, stylistics and pragmatics, focusing on political and academic discourse, and is currently involved in the research project *Coherence and Cohesion in English Discourse*, the aim of which is to conceptualize coherence and cohesion as constitutive components of human communication and to explore features and strategies enhancing the perception of coherence and cohesion in different genres of spoken and written discourse. She co-edits the academic journal *Discourse and Interaction*.

Address: Mgr. Olga Dontcheva-Navratilova, Ph.D., Department of English Language and Literature, Faculty of Education, Masaryk University, Poříčí 9, Brno 603 00, Czech Republic. [e-mail: navratilova@ped.muni.cz]