

# LEXICAL BUNDLES INDICATING AUTHORIAL PRESENCE: A CROSS-CULTURAL ANALYSIS OF NOVICE CZECH AND GERMAN WRITERS' ACADEMIC DISCOURSE

*Olga Dontcheva-Navratilova*

## Abstract

With the widespread use of English as the *lingua franca* of academia, there is a growing need of research into how non-native speakers striving to be socialized in target academic discourse communities deal with variation in meaning and organization of academic texts across fields, languages and cultures. An important indicator of competent linguistic production is the mastering of the register- and genre-specific formulaic expressions termed lexical bundles, which are defined as sequences of three or more words with frequent co-occurrence in a particular context (Biber et al. 1999). While recent studies have addressed disciplinary and novice-expert differences in the use of lexical bundles, cross-cultural variation in bundle use remains underexplored. This paper investigates lexical bundles indicating authorial presence in a specialized corpus of Master's degree theses from the fields of linguistics and methodology written by German and Czech university students. The aim of the study is to compare how novice Czech and German authors use lexical bundles indicating authorial presence, to consider whether and to what extent the novice writers have adapted their writing style to the conventions of Anglo-American academic writing, and to discuss the role of the L1 academic literacy tradition and instructions received in writing courses for the modelling of novice writers' academic discourse. The analysis shows that the variety and frequency of interpersonal bundles in Czech and German novice writers' discourse do not approximate to the standard of published academic texts in English. The findings also indicate that while the considerable similarities in the way Czech and German novice writers use the target structures for constructing authorial presence reflect their common roots in the Central European tradition of academic discourse, the divergences may be attributed to a difference in the degree of adaptation to Anglo-American writing conventions.

## Keywords

authorial presence, cross-cultural variation, lexical bundles, novice academic discourse

## 1 Introduction

It is now a universally acknowledged truth that English has become the *lingua franca* of the global academic world. One of the consequences of this widespread use of English in academia is that the majority of its users are non-native speakers. Yet until recently most descriptive and pedagogical studies of academic English have taken the educated native speaker as the model of good and fluent language performance and the Anglo-American tradition of academic writing as the prevailing discourse convention, thus creating a centre-periphery

tension and relegating non-native-speakers to marginal participation in the activities of the academic discourse community (cf. Flowerdew J. 2000, Salager-Meyer 2008). However, in the last decade numerous researchers and educational practitioners have problematized the role of the native speaker as a model and questioned the practice of imposing the Anglo-American tradition of academic writing on intercultural communication (e.g. Flowerdew J. 2008, Jenkins 2009, Mauranen 2009), pointing out that “there are no native speakers of *academic English*” (Mauranen, Hynninen & Ranta 2010: 184) and that all novices, irrespective of their first language, have to undergo a secondary socialization when knocking at the door of academia. This has called for research into how novice non-native speakers acquire the norms of interaction of target academic discourse communities, how they adapt to the existing variation in meaning and organization of academic texts across fields, languages and cultures, and what communicative strategies they need to use in order to be accepted by the institutional gatekeepers, such as examination boards, journal editors and peer-reviewers.

Maybe the best opportunity for integration of novice non-native speakers of English into academia is provided by universities, where their socialization is facilitated by instruction and involvement in the activities of the target discourse community (Flowerdew J. 2000). As the master’s degree thesis is “the most sustained and complex piece of academic writing” (Swales 2004: 99) that all master’s students undertake, it appears that it is the most appropriate kind of text for an exploration of the academic writing habits of non-native speakers standing on the threshold of academia. From a cross-cultural perspective this paper explores diploma theses written by novice German and Czech speakers of academic English in order to find out how they use lexical bundles to construe authorial presence – the degree of visibility and authoritativeness writers are prepared to project in their texts for personal support of their statements when expressing their attitudes, judgements and assessments. The choice of lexical bundles as the target structures under analysis draws on an understanding of the acquisition of conventionalized expressions used by a target discourse community as one of the indicators of the communicative competence achieved by learners during their socialization in an academic setting (Cortes 2004). Since the construal of authorial presence has been identified as one of the aspects of academic writing marked by most prominent cross-cultural variation, the main purpose of this investigation is to explore how novice Czech and German writers use lexical bundles to convey interpersonal meanings and to organize their texts, while considering whether and to what extent they have adapted their writing style to the Anglo-American academic convention.

## 2 Cross-cultural differences in the construal of authorial presence

Recent insights into academic writing in English (e.g. Chamonikolasová 2005, Čmejrková & Daneš 1997, Duszak 1994, Mur-Dueñas 2007, Povolná 2010) have shown that the international academic community “shares field interests, yet falls short of common language and discourse properties” (Duszak 1997: 20), as there is considerable variation in text characteristics across fields, language and cultures. As a result, a novice non-native speaker of academic English has to face the challenging task of secondary socialization to perform a new public role within a new discourse community controlled by a literacy practice shaped by different social and cultural conventions. While striving to achieve this, novice non-native speakers, especially those from less central discourse communities (cf. Kachru 2001), often experience identity instability (cf. Ivanič 1998, LeCourt 2004), which may be reflected in the construal of authorial presence in their academic discourse, since they may be reluctant to take on the identity of a member of the target community, i.e. “the identity of a person with authority” (Ivanič 1998: 88).

When undertaking a cross-cultural analysis of academic discourse, it is essential to begin by describing and comparing the literacy traditions that the texts under analysis represent. In this case these are the Anglo-American and the Central European traditions of academic writing. Belonging to the same Central European literacy tradition, Czech and German academic writing are expected to share a lot of distinctive features; the similarity is supposed to be further enhanced by the fact that Czech academic writing was historically under the influence of the Teutonic intellectual style (Galtung 1981, as quoted in Clyne 1987). However, while according to Clyne (1987: 233) English texts by German scholars tend to contain the same cultural discourse patterns as German texts, Čmejrková and Daneš (1997: 42) claim that Czech academic writing has recently been profoundly affected by the spread of English academic norms. This may reflect the more peripheral status of Czech academic discourse as well as a historically motivated tendency to adapt to a dominant culture.

As evidenced by previous research (e.g. Chamonikolasová 2005, Clyne 1987, Duszak 1997, Kreutz & Harres 1997, Stašková 2005), the Anglo-American and Central European academic discourses differ considerably in the way they approach writer-reader interaction and discourse organization (cf. Table 1). Some of the differences seem to stem from the size of the respective discourse communities and the impact of this on solidarity and power relations among their members (Čmejrková & Daneš 1997). Thus patterns of interaction in small communities (such as the Czech one) tend to be marked by symbiosis

and avoidance of tension, as the considerable amount of common knowledge and shared methodological principles allow the members of the discourse community to make their rather monologic texts more implicit and less structured, while leaving the responsibility for understanding the text to the reader. On the other hand, members of the highly competitive English-speaking academic discourse community strive to find a research space in a research territory densely packed with occupied “niches” (Duszak 1997). When trying to persuade a depersonalized readership to accept their views, researchers working within the Anglo-American academic tradition typically adopt a more reader-friendly attitude associated with a higher level of interactivity and explicit discourse organization, as it is the writer who takes responsibility for making the text understandable. As Čmejrková and Daneš (1997: 57) point out, it is precisely these features that make the present-day Anglo-American norm of academic writing particularly suitable for the purposes of cross-cultural communication within the international discourse community, in which the writer cannot assume a high degree of shared knowledge, experience and discourse conventions.

Anglo-American academic writing	Central European academic writing
<ul style="list-style-type: none"> <li>– competitive large discourse communities</li> <li>– interactive, dialogic</li> <li>– negotiation of meaning</li> <li>– reader-oriented</li> <li>– marked authorial presence</li> <li>– strict discourse norms</li> <li>– explicit discourse organization</li> </ul>	<ul style="list-style-type: none"> <li>– small discourse communities avoiding tension</li> <li>– low-interactive, monologic</li> <li>– conceptual and terminological clarity</li> <li>– writer-oriented</li> <li>– backgrounded authorial presence</li> <li>– absence of strict discourse norms</li> <li>– low on explicit discourse organization</li> </ul>

**Table 1: The Central European vs. the Anglo-American academic writing tradition**

The dialogic character of the Anglo-American academic discourse favours a marked authorial presence helping writers to negotiate claims and debate views with the implied audience, while facilitating the reader’s path through the text by previewing the content of upcoming discourse, staging and signposting. This interpersonal dimension of discourse is conveyed by attitudinal markers (e.g. hedges, boosters, personal intrusions) modifying the force of the argument and appealing to the reader in seeking agreement with the viewpoint advanced by the author, and discourse organizers indicating intratextual connections and logical relations holding in the discourse. Unlike with the Anglo-American tradition, the focus of Central European academic writing is on conceptual and terminological clarity rather than persuasion and discourse organization and thus authorial presence tends to be backgrounded. However, previous studies have highlighted

some variation within the Central European tradition: while German academic discourse, considered to be oriented towards establishing authority in the discipline, uses very few mitigating devices (Kreutz & Harres 1997: 181), Czech academic writing is characterized by a lesser degree of assertiveness expressed by the use of face-saving devices such as hedging predicates, participles and adverbs (Čmejrková & Daneš 1997: 49).

Obviously, the various academic writing traditions are transmitted through the educational system and academic style manuals targeted at a readership of novice writers. As a recent survey of printed academic style manuals (Bennett 2009: 43) has shown, there is “a remarkable consensus as regards general principles, methods of textual construction, and the kinds of grammatical and lexical features to be used”, which indicates the persistence of the so-called scientific paradigm in academic English related to “clarity, economy, rational argument supported by evidence, caution and restraint” (Bennett 2009: 52). As far as authorial presence is concerned, however, Bennett mentions some disagreement about the use of personal or impersonal forms, which is also highlighted by Flowerdew L. (2012), who notes that “contrary to advice given in some style guides to maintain an objective, impersonal style, the pronoun system is exploited by writers of RAs [research articles] for maintaining the writer-reader relationship and allowing the writer an authorial voice”. Within the Central European tradition, the tendency to background authorial presence concurs with the use of impersonal structures, and, in the case of personal structures, with the use of first person plural forms (Chamonikolasová 2005: 82). It is therefore evident that besides the variation in the advice provided by different style manuals there is some difference in the structures typically used for the construal of authorial presence in the Anglo-American and Central European academic discourses. That is why the intention of the present investigation is to explore how the L1 tradition, the advice provided by style manuals, and instructions received in academic writing courses have affected the academic discourse of Czech and German novice writers.

### **3 Lexical bundles**

The paramount importance of conventionalized expressions as an essential aspect of the shared knowledge of a professional discourse community and an indicator of competent use of language in a particular context has been shown conclusively by a large number of corpus-based investigations of academic English (cf. Cortes 2002, Hyland 2008, Nattinger & DeCarrico 1992, Simpson 2004, Wray 2000, 2002). Adopting a frequency-driven approach, Biber et al. (1999: 990) have introduced the term ‘lexical bundles’ to refer to “recurrent

expressions, regardless of their idiomaticity, and regardless of their structural status”. In subsequent studies lexical bundles have been defined as the most frequent recurrent sequences of three and more words in a register or genre (Biber & Barbieri 2007: 264) and are thus seen as indicators of professional proficiency rather than native-like language fluency. The frequency cut-off applied in different studies varies from 40 per million words (e.g. Biber & Barbieri 2007) to ten per million words (e.g. Simpson-Vlach & Ellis 2010). The distinctive features of lexical bundles which distinguish them from other kinds of formulaic expressions are their frequency of occurrence, non-idiomaticity, lack of perceptual salience and structural incompleteness (Biber & Barbieri 2007: 269). While it is these very features that seem to have kept bundles outside the scope of the expressions traditionally included in foreign language teaching curricula, there is now a growing awareness of the necessity of including overt teaching of lexical bundles in academic writing courses, which is evidenced, for instance, by Simpson-Vlach and Ellis’s (2010) Academic Formulas List for pedagogical purposes.

Lexical bundles can be categorized according to their structural and functional characteristics. Despite minor differences in the structural classifications suggested by previous research (e.g. Biber 2006, Biber & Barbieri 2007, Hyland 2008), bundles can be regarded as pertaining to four categories, namely, (i) verb phrase components, including passive voice, anticipatory *it* structures and dependent clause fragments (*is assumed to be*, *it is possible that*), (ii) noun phrase elements followed by a part of a modifier (*the end of the*, *the extent to which*), (iii) prepositional phrases followed by prepositional or clausal elements (*at the end of*, *of the things that*), and (iv) formulaic structures (e.g. *have a nice day*). The personal and impersonal bundles indicating authorial presence analysed in this study are subsumed under the first structural category – verb phrase components.

From a functional perspective, lexical bundles can be regarded as “a kind of pragmatic ‘head’ for larger phrases or clauses, where they function as discourse frames for the expression of new information” (Biber & Barbieri 2007: 270). Drawing on the functional taxonomies suggested by Biber, Conrad and Cortes (2004), Cortes (2004), Biber and Barbieri (2007), Hyland (2008) and Simpson-Vlach and Ellis (2010), bundles are seen here as performing three major discourse functions – referential, discourse organizing and attitudinal – which, as Cortes (2004: 401) points out, reflect ideational, textual and interpersonal meanings (Halliday 1985), respectively.

1. Referential bundles 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 – *a little bit of, the structure of the*
- c) topic-specific bundles – *in the curricula of, the interpretation of the*
2. Discourse organizers 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*
    - resultative signals – *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*
  - c) framing bundles (focusing, limiting conditions on arguments) – *in the case of, with the exception of*
3. Attitudinal bundles convey interpersonal meanings including:
  - a) stance bundles (express the evaluations and attitudes of the writer) – *it is (im)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*

Due to variation in their pragmatic meanings, which is enhanced by their structural incompleteness, some lexical bundles can have multiple functions even within a single occurrence (Biber, Conrad & Cortes 2004: 383, Nesi & Basturkmen 2009: 32). This has resulted in some discrepancies among the existing functional classifications, which seem to reflect variation in the size and composition of the corpora used by the researchers; nevertheless, it can be assumed that bundles typically fulfil one main function, which may be register-, discipline- or genre-specific. Thus 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).

By organizing discourse into a lesser number of larger meaningful units all three functional types of bundles can be seen as discourse signals which facilitate accurate understanding and ease fluent discourse production by contributing to the perception of coherence in discourse: referential bundles promote topic continuity, attitudinal bundles help build up a consistent authorial voice, while discourse organizers help the reader to establish relations between adjacent and distant stretches of discourse. The latter two categories, which convey interpersonal meanings and contribute to construal of authorial presence, are in the focus of the present study.

## 4 Material and methodology

This investigation of lexical bundles indicating authorial presence in novice non-native speaker academic discourse explores the occurrence of highly frequent bundles identified by previous studies (e.g. Biber 2006, Biber & Barbieri 2007, Cortes 2002, Hyland 2008, Simpson-Vlach & Ellis 2010) in diploma theses written by Czech and German students of English. The aim of the quantitative analysis is to find similarities and differences in the use of target personal and impersonal attitudinal and discourse organizing bundles in Czech and German diploma theses. The results of the analysis serve as the basis for a discussion of issues of cross-cultural variation, the influence of L1 traditions of academic literacy, and advice provided by style manuals, academic writing instructors and diploma thesis supervisors.

The investigation is carried out on two small specialized corpora representing Czech and German novice academic writing. Although small corpora have been criticized for their limited size and representativeness, and the generalizability of their results has been questioned, there are some strong arguments for their use in comparative studies of academic and professional discourse, especially when focusing on particular language features in a given type of text. As Flowerdew L. (2004: 18) points out, “specialized corpora allow for more top-down, qualitative, contextually-informed analyses than those carried out using general corpora”. The material used in this research comprises a Czech corpus consisting of ten Master’s degree theses (5 in the field of linguistics and 5 in the field of methodology) written by students of English at Masaryk University, and a German corpus including seven Master’s degree theses (4 in the field of linguistics and 3 in the field of methodology) written by students of English at Chemnitz University; all the theses were written in the period 2005 to 2009 and they represent the top 20 per cent of the results achieved in this period. While the Czech and the German corpora differ in the number of theses included, they are of the same size in terms of wordcount. Thus the Czech corpus, which is further subdivided into a linguistics theses sub-corpus and a methodology theses sub-corpus (each of 88,000 words) has a total wordcount of 176,000 words, and the German corpus, which is also subdivided into a linguistics part (91,000 words) and a methodology part (85,000 words), also amounts to 176,000 words, i.e. the total size of the material is 352,000 words. The corpus was cleaned to eliminate the text of block quotes and long examples; however, the text of integral citations and integral examples was not deleted in order to preserve the coherence of the texts. The cleaned corpus was searched for the target bundles using Antconc, a concordance programme; the raw data were then normalized to frequencies per million words to allow for comparison with data reported by previous studies.



Following Nattinger and DeCarrico's (1992) approach, the set of target expressions was selected prior to the study; the choice of bundles considered representative of similar genres and disciplines was based on the results of previous research, namely Hyland's (2008) research into bundles used in research articles and doctoral and Master's theses in electrical engineering, biology, applied linguistics and business studies, Cortes's (2004, 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. This allows an assessment of the extent to which novice writers have acquired register-specific bundles, which makes this approach particularly appropriate for the analysis of novice non-native speakers' academic discourse. The expressions under analysis are mainly 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-word or more-word bundles. It should be noted, however, that as the rate of discourse organizers and some personal structures was found to be rather low, some target bundles allow for some variation. Because of the lower frequency of occurrence of academic formulas in novice non-native speaker discourse, the frequency cut-off for considering that an expression has reached a relevant frequency 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 the majority of the diploma theses (in all but one in each sub-corpus) to be considered as yielding significant frequency rates.

The 26 attitudinal and discourse organizing bundles (target bundles) analyzed in this investigation (cf. Table 2) were chosen to represent the most frequent expressions used in the field of the humanities; despite some differences in the results reported in previous studies, the frequency of occurrence of these bundles in expert academic discourse is assumed to exceed 50 per million words. Since novice non-native writers are likely to acquire general academic vocabulary first during their university studies, the target bundles were selected so as to represent core academic bundles, as classified in Simpson-Vlach and Ellis's (2010) Academic Formulas List.

The target bundles listed in Table 2 below are grouped according to form (personal vs. impersonal structures) and function (attitudinal bundles and discourse organizers). Since the selection is frequency-based, it does not represent evenly all the structural and functional types of lexical bundles and

cannot illustrate all the rhetorical strategies used by the students in the construal of authorial presence; yet it is assumed that the set of target bundles can outline significant differences in Czech and German novice writers' academic discourse. The most numerous type of bundles included in this investigation is the impersonal attitudinal bundle comprising an anticipatory *it* followed by an adjective phrase or a passive construction (13 structures), while personal attitudinal bundles are represented by five structures consisting of a personal pronoun followed by a verb phrase. The smaller group of discourse organizers is composed of four personal and four impersonal bundles including a personal pronoun and/or a text/research noun followed by a discourse/research verb. It should be noted that the results of the quantitative analysis classify multifunctional bundles according to what is considered to be their main function; however, in some cases this may affect the results, as not only can bundles perform different functions in different contexts, but they can have more than one function within a single occurrence.

Target bundles	Personal structures	Impersonal structures
<b>Attitudinal bundles</b>	I/we assume that the/this I/we believe that the/this I/we consider that the/this I/we think that the/this I/we would like to (stress/note/ mention)	<i>it is important to</i> <i>it is clear that</i> <i>it is possible to</i> <i>it is difficult to</i> <i>it is impossible to</i> <i>it is interesting to</i> <i>it is obvious that</i> <i>it is likely that</i> <i>it is necessary to</i> <i>it should be noted</i> <i>it could be argued</i> <i>it is assumed that</i> <i>it can be seen</i>
<b>Discourse organizers</b>	<i>I am going to/will focus on/study</i> <i>in this chapter/thesis/work/paper</i> I/we <i>my results prove/show/indicate</i> <i>(that)</i> <i>I/we have shown/proved/found that</i>	<i>this/the thesis/paper focuses on/</i> <i>studies</i> <i>this/the thesis/paper will study/</i> <i>focus on</i> <i>the results suggest/show/indicate</i> <i>that</i> <i>it has been/was found/proved/</i> <i>shown that</i>

**Table 2: Target bundles**

## 5 Comparison of the use of the target bundles in the Czech and the German corpora

The comparative analysis first focuses on the rate of occurrence of the target bundles in the Czech and the German corpora and relates it to the standard rate reported for academic discourse. The frequency information summarized in Table 3 shows that there are considerable differences in the frequency of use of the target bundles between the Czech and the German corpora. The overall density of target bundles is considerably higher in the Czech corpus, which seems to reflect instructions received in academic writing courses, as well as a more marked tendency in Czech academic writing to adapt to the conventions of the dominant Anglo-American writing standard. Of course, it should be acknowledged that apart from indicating cross-cultural variation, the discrepancies in the rate of the target bundles may stem from the authors' preference towards the use of other structures.

Target bundles	German		Czech	
	Raw No.	% of total words in target bundles	Raw No.	% of total words in target bundles
<b>Attitudinal bundles (total)</b>	<b>90</b>	<b>0.05</b>	<b>186</b>	<b>0.10</b>
Impersonal	63	0.017	143	0.08
Personal	27	0.015	43	0.02
<b>Discourse organizers (total)</b>	<b>20</b>	<b>0.01</b>	<b>17</b>	<b>0.009</b>
Impersonal	17	0.009	9	0.005
Personal	3	0.001	8	0.004
<b>Total target bundles</b>	<b>110</b>	<b>0.06</b>	<b>203</b>	<b>0.11</b>

**Table 3: Frequency of target bundles in German and Czech corpora**

The results of the quantitative analysis also show that the frequency of use of the target bundles by Czech and German thesis writers is lower than that typical of academic discourse. This is evidenced by the low total percentage of words in bundles – 0.06 for the German corpus and 0.11 for the Czech corpus – when compared to the data reported in Hyland's (2008) study, where the proportion of texts comprised of words in bundles was found to be 1.9 per cent. The difference is so significant that it cannot be attributed only to the restricted set of structures included in the research and the exclusion of referential bundles, regarded as the most prominent type of bundles in academic prose (Biber 2006). This lower bundle density is in consonance with previous studies on novice academic

writers' discourse (e.g. Cortes 2002, 2004) and strengthens the argument for overt instruction in discipline- and genre-specific bundles in academic writing courses, as covert teaching through exposure to their frequent occurrence in published academic texts does not seem to lead to successful acquisition of these expressions by university students.

In both corpora, the target attitudinal bundles show a higher frequency of occurrence, which reflects the argumentative character of humanities, while the prevalence of impersonal structures is in consonance with the advice provided by most academic writing style manuals and the tendency in Central European academic discourse to background authorial presence. Although most style manuals emphasize the need of explicit signposting for achieving better coherence, the rate of the target discourse organizers in the corpora is low. Apart from indicating that novice writers have not yet developed sufficiently their rhetorical skills, this seems to be grounded in the more topic-oriented and writer-centred Central European tradition of academic writing, which tends to be structured less transparently and to use sparingly staging and signposting devices guiding the reader through the text.

Despite the similarity in these general tendencies there are differences in the relative rate of the structural and functional types of target bundles present in the Czech and the German theses. The Czech novice authors use twice as many attitudinal bundles as the German novice authors. This can be attributed to the influence of the Czech academic writing tradition, which is characterized by a lesser degree of assertiveness conveyed by the use of epistemic stance bundles (e.g. *I would like to, it can be argued*) and hedges as part of discourse organizers (*the results suggest/indicate*), as well as to the efforts of Czech novice writers to prepare the ground for an acceptance of their claims by the use of interactive structures expressing judgements of necessity/importance (*it is necessary to, it is important to*). Following the general orientation of German academic discourse towards establishing authority in the discipline, German novice writers use fewer mitigating devices and prefer to build up their arguments using hypothetical *if*-structures typical of writing in the natural sciences and empirical social sciences. The density of discourse organizers is slightly higher in the German corpus; what is striking, however, is the marked preference of German novice writers towards the use of impersonal structures, while Czech authors use approximately the same number of personal and impersonal discourse organizers.

As shown in Table 4, there are also differences between the rates of target bundles in the linguistics and methodology theses, pointing to variation in what the writers are trying to achieve in their texts. It is noteworthy that the German corpus shows no considerable variation in the density of target bundles, except

for the higher rate of impersonal discourse organizers in the methodology theses. This consistency of German novice discourse across disciplines seems to be the result of instructions provided in academic writing courses at Chemnitz University and the influence of the strong German academic writing tradition, the discourse patterns of which tend to be transposed into texts written in English (Clyne 1987). In contrast, the preferences of Czech novice writers of linguistics and methodology theses differ considerably, reflecting the change that Czech academic discourse is undergoing under the influence of English academic writing conventions and discipline-specific advice provided by the diploma thesis supervisor. The high rate of impersonal attitudinal bundles in linguistics theses, which typically deal with the interpretation of language data, may be regarded as an indicator of the reluctance of the authors to adopt an authoritative authorial voice. The lower incidence of attitudinal bundles in Czech methodology theses seems to reflect their more descriptive orientation.

Target bundles	Linguistics theses		Methodology theses	
	German	Czech	German	Czech
<b>Attitudinal bundles (total)</b>	<b>45</b>	<b>130</b>	<b>45</b>	<b>56</b>
Impersonal	29	101	34	42
Personal	16	29	11	14
<b>Discourse organizers (total)</b>	<b>7</b>	<b>10</b>	<b>13</b>	<b>7</b>
Impersonal	6	4	11	5
Personal	1	6	2	2
Total target bundles	<b>52</b>	<b>140</b>	<b>58</b>	<b>63</b>

**Table 4: Comparison of frequency of target bundles in linguistics and methodology theses**

Apart from the differences in the rate of target bundles, the analysis has shown that Czech and German novice writers differ in the repertoire of bundles that they use. Table 5 lists the 17 out of 26 target bundles that have reached a relevant frequency of 20 occurrences per million words in at least one sub-corpus. The only target bundles which have attained relevant frequency in all sub-corpora are two impersonal attitudinal bundles using the anticipatory *it*-structure: the evaluative expression *it is important to (note/mention)* conveying the writer's judgment on the importance of a point made in their argumentation, and the interactional structure *it can be argued/noted/stated*, by which the writers indicate the non-factual status of the proposition and thus involve the reader in the argumentation. The pragmatic functions of evaluative stance bundles are basically identical in both corpora, although Czech authors use a wider repertoire

of these expressions (*it is important to, it is possible to, it is necessary to, it is obvious that*). In the case of personal stance bundles, there is some difference in the choice of stance verbs controlling *that*-clauses – *assume* and *consider*, which function as hedges expressing a reluctance to convey complete commitment to a claim, occur in the German corpus, while the Czech authors prefer the markers of greater writer involvement *believe* and *think*. Similarly to Hyland's (2008) findings, the personal epistemic stance bundle *I would like to* has the highest rate of occurrence; in both the German and Czech theses corpora it 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. It is interesting to note that linguistics theses use the widest repertoire of bundles, which may be tentatively explained by a higher degree of language awareness on the part of the student-writers. In contrast, the authors of methodology theses, especially in the German corpus, seem to favour the use of a limited number of highly frequent structures. Such a frequent use of a specific bundle may be seen as a form of overgeneralization (Cortes 2004: 412), i.e. a stage in the process of acquisition of academic discourse, during which novice writers use bundles with a high rate of occurrence before learning how to use them in an appropriate and balanced way.

Target bundles	German corpus				Czech corpus			
	Linguistics		Methodology		Linguistics		Methodology	
	Raw No.	Norm. rate	Raw No.	Norm. rate	Raw No.	Norm. rate	Raw No.	Norm. rate
<b>Attitudinal bundles (impersonal)</b>								
<i>It is important to (note/mention)</i>	8	<b>88</b>	10	<b>118</b>	9	102	15	170
<i>It is possible to</i>	1	<b>11</b>	2	<b>24</b>	31	352	6	68
<i>It is impossible to</i>	6	<b>66</b>	0	<b>0</b>	3	34	1	11
<i>It is difficult to</i>	0	<b>0</b>	0	<b>0</b>	4	45	0	0
<i>It is obvious that</i>	2	<b>22</b>	0	<b>0</b>	5	57	6	68
<i>It is necessary to</i>	3	<b>33</b>	0	<b>0</b>	14	<b>159</b>	6	<b>68</b>
<i>It can be argued/noted/stated</i>	3	<b>33</b>	16	<b>188</b>	28	318	5	57
<i>It should be noted</i>	1	<b>11</b>	3	<b>35</b>	3	<b>34</b>	0	<b>0</b>
<b>Attitudinal bundles (personal)</b>								
<i>I assume that (the/this)</i>	7	<b>77</b>	1	<b>12</b>	1	11	0	0

**LEXICAL BUNDLES INDICATING AUTHORIAL PRESENCE:  
A CROSS-CULTURAL ANALYSIS OF NOVICE CZECH AND GERMAN WRITERS' ACADEMIC DISCOURSE**

<i>I consider (that) the/this</i>	8	<b>88</b>	0	<b>0</b>	0	0	3	34
<i>I would like to</i>	0	<b>0</b>	9	106	23	261	5	57
<i>I think that the/this</i>	0	<b>0</b>	0	<b>0</b>	0	0	4	45
<i>I believe that the/this</i>	1	<b>11</b>	1	<b>12</b>	5	57	2	23
<b>Discourse organizers (impersonal)</b>								
<i>This thesis/paper focuses on/studies</i>	3	<b>33</b>	2	<b>24</b>	0	<b>0</b>	0	<b>0</b>
<i>The results suggest/indicate that</i>	1	<b>11</b>	4	<b>47</b>	4	45	5	57
<i>It has been proved/shown that</i>	1	<b>11</b>	4	<b>47</b>	0	0	0	0
<b>Discourse organizers (personal)</b>								
<i>In this thesis, I/we will/shall</i>	0	<b>0</b>	0	<b>0</b>	3	<b>34</b>	1	<b>11</b>

**Table 5: Comparison of frequency of relevant bundles in the German and Czech corpora**

As to discourse organizers, the resultative logical relations bundle *The results suggest/indicate that*, which also has a hedging function, shows relevant frequency in both corpora, although its rate in the Czech corpus is higher, which is in consonance with the tendency towards a higher degree of tentativeness in Czech academic discourse. The second resultative bundle *It has been proved/shown that* occurs only in the German theses and indicates a higher degree of assertiveness. The use of the intratextual reference bundles is highly indicative of differences in the preferences of Czech and German novice writers: while German authors use the impersonal bundle *this paper focuses on*, Czech authors prefer the personal bundle *In this thesis I/we will/shall*. It should be noted that although this use of *I* foregrounds the author as an ‘architect’ of the thesis, it represents the less authoritative end of the cline showing the degree of power wielded by the authorial presence, as described by Tang and John (1999).

## 6 Conclusions

The main purpose of this investigation was to compare how novice Czech and German writers use lexical bundles indicating authorial presence, while considering whether and to what extent they have adapted their writing style to the conventions of Anglo-American academic discourse. The findings have shown that in agreement with results reported by previous studies into novice academic discourse the variety and frequency of interpersonal bundles in Czech

and German novice writers' discourse do not approximate to the standard of published academic texts in English. Apart from the insufficiently developed rhetorical skills of the authors, this can be explained by the difference between L1 and Anglo-American academic discourse conventions. The comparative analysis of the target bundles in novice Czech and German theses has indicated that owing to their common roots in the Central European tradition of academic discourse there are considerable similarities in the way novice writers use the target structures for constructing authorial presence. The differences between Czech and German novice academic discourse may be attributed to a different degree of adaptation of Czech and German academic writing in English to Anglo-American writing conventions and to advice received from supervisors and teachers of academic writing.

In conclusion, it can be stated that the results of this study show that cross-cultural analysis of bundles indicating authorial presence may contribute to a better understanding of the reasons for the existing variation in academic discourse conventions across languages and cultures and thus help novice non-native writers to be socialized into the global academic discourse community. Further research into lexical bundles and various other devices for the construal of authorial presence should concentrate on the extent to which cross-cultural variation has affected the international academic norm.

## References

- Bennett, K. (2009) 'English academic style manuals: A survey.' *Journal of English for Academic Purposes* 8/1, 43-54.
- Biber, D. (2006) *University Language: A Corpus Study of Spoken and Written Registers*. Amsterdam and Philadelphia: John Benjamins.
- Biber, D. and Barbieri, F. (2007) 'Lexical bundles in university spoken and written registers.' *English for Specific Purposes* 26/3, 263-286.
- Biber, D., Conrad, S. and Cortes, V. (2004) "'If you look at...': Lexical bundles in university teaching and textbooks.' *Applied Linguistics* 25/3, 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Chamonikolasová, J. (2005) 'Comparing the structures of academic texts written in English and Czech.' In: Huttová, M. et al. (eds) *Slovak Studies in English 1*. Bratislava: Univerzita Komenského. 77-84.
- Clyne, M. (1987) 'Cultural differences in the organisation of academic texts.' *Journal of Pragmatics* 11, 211-247.
- Čmejrková, S. and Daneš, F. (1997) 'Academic writing and cultural identity: The case of Czech academic writing.' In: Duzsak, A. (ed.) *Culture and Styles of Academic Discourse*. Berlin and New York: Mouton de Gruyter. 40-62.
- Cortes, V. (2002) 'Lexical bundles in Freshman composition.' In: Reppen, R., Fitzmaurice, S. and Biber, D. (eds) *Using Corpora to Explore Linguistic Variation*. Amsterdam and Philadelphia: John Benjamins. 131-146.



- Cortes, V. (2004) 'Lexical bundles in published and student disciplinary writing: Examples from history and biology.' *English for Specific Purposes* 23/4, 397-423.
- Duszak, A. (1997) 'Cross-cultural academic communication: A discourse community view.' In: Duszak, A. (ed.) *Culture and Styles of Academic Discourse*. Berlin and New York: Mouton de Gruyter. 11-39.
- Duszak, A. (1994) 'Academic discourse and intellectual styles.' *Journal of Pragmatics* 21/3, 291-313.
- Flowerdew, J. (2000) 'Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar.' *TESOL Quarterly*, Vol. 34/1, 127-150. Retrieved from <http://www.jstor.org/stable/3588099>.
- Flowerdew, J. (2008) 'Scholarly writers who use English as an additional language: What can Goffman's "Stigma" tell us?' *Journal of English for Academic Purposes* 7/2, 77-86.
- Flowerdew, L. (2012) 'Grammar and the research article.' In: Chapelle, C. (ed.) *Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- Flowerdew, L. (2004) 'The argument for using English specialized corpora to understand academic and professional language.' In: Connor, U. and Upton, T. (eds) *Discourse in the Professions*. Amsterdam and Philadelphia: John Benjamins. 11-33.
- Halliday, M. A. K. (1985) *An Introduction to Functional Grammar*. London: Edward Arnold.
- Hyland, K. (2008) 'As can be seen: Lexical bundles and disciplinary variation.' *English for Specific Purposes* 27/1, 4-21.
- Ivanič, R. (1998) *Writing and Identity. The Discoursal Construction of Identity in Academic Writing*. Amsterdam and Philadelphia: John Benjamins.
- Jenkins, J. (2009) '(Un)pleasant? (In)correct? (Un)intelligible? ELF speakers' perceptions of their accents.' In: Mauranen, A. and Ranta, E. (eds) *English as a Lingua Franca: Studies and Findings*. Newcastle upon Tyne: Cambridge Scholars Publishing. 10-36.
- Kachru, B. (2001) 'World Englishes.' In: Mesthrie, R. (ed.) *Concise Encyclopedia of Sociolinguistics*. New York: Elsevier. 519-524.
- Kreutz, H. and Harres, A. (1997) 'Some observations on the distribution and function of hedging in German and English academic writing.' In: Duszak, A. (ed.) *Culture and Styles of Academic Discourse*. Berlin and New York: Mouton de Gruyter. 181-201.
- LeCourt, D. (2004) *Identity Matters: Schooling the Student Body in Academic Discourse*. Albany: State University of New York Press.
- Mauranen, A. (2009) 'Introduction.' In: Mauranen, A. and Ranta, E. (eds) *English as a Lingua Franca: Studies and Findings*. Newcastle upon Tyne: Cambridge Scholars Publishing. 1-7.
- Mauranen, A., Hynninen, N. and Ranta, E. (2010) 'English as an academic lingua franca: The ELFA project.' *English for Specific Purposes* 29/3, 183-190.
- Mur-Dueñas, P. (2007) 'I/we focus on...: A cross-cultural analysis of self-mentions in business management research articles.' *Journal of English for Academic Purposes* 6/2, 143-162.
- Nattinger, J. and DeCarrico, J. (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesi, H. and Basturkmen, H. (2009) 'Lexical bundles and discourse signalling in academic lectures.' In: Flowerdew, J. and Mahlberg, M. (eds) *Lexical Cohesion and Corpus Linguistics*. Amsterdam and Philadelphia: John Benjamins. 23-43.
- Povolná, R. (2010) 'Exploring sequential relations in learner discourse.' In: Ciuk, A. and Molek-Kozakowska, K. (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.
- Salager-Meyer, F. (2008) 'Scientific publishing in developing countries: Challenges for the future.' *Journal of English for Academic Purposes* 7/2, 121-132.
- Simpson, R. (2004) 'Stylistic features of academic speech: The role of formulaic expressions.' In: Connor, U. and Upton, T. (eds) *Discourse in the Professions. Perspectives from Corpus Linguistics.* Amsterdam and Philadelphia: John Benjamins. 37-64.
- Simpson-Vlach, R. and Ellis, N. (2010) 'An academic formulas list: New methods in phraseology research.' *Applied Linguistics* 31/4, 463-512.
- Stašková, J. (2005) 'Options of identity: Authorial presence in research articles abstracts.' In: Huttová, M. et al. (eds.) *Slovak Studies in English 1.* Bratislava: Univerzita Komenského. 201-207.
- Swales, J. (2004) *Research Genres. Exploration and Applications.* Cambridge: Cambridge University Press.
- Tang, R. and John, S. (1999) 'The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun.' *English for Specific Purposes* 18/1, 23-39.
- Wray, Alison (2000) 'Formulaic sequences in second language teaching: Principle and practice.' *Applied Linguistics* 21/4, 487-489.
- Wray, A. (2002) *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.