LEXICAL BUNDLES IN ACADEMIC ARTICLES FROM ARTS DISCIPLINES

Hana Atcheson

Abstract

This paper studies lexical bundles in academic articles in the fields of photography, product/industrial design and audiovisual arts disciplines. The comparative corpus study presents a quantitative and qualitative analysis of formal and functional types of 4-grams used in published articles by expert authors in the three arts disciplines. The analysis was based on Biber et al.'s (1999) formal taxonomy and Hyland's (2008a, 2008b) functional taxonomy of lexical bundles. The results show that both photography and audiovisual arts disciplines have similar utilisation of formal and functional types of bundles, prioritising content-oriented rhetorical functions. However, the discipline of product/industrial design shows a stronger preference for participant-oriented functions that utilise a distancing, hedging type of bundle. The findings help teachers of academic writing to identify the objective language needs of PhD students for writing in these disciplines. Writing instruction can be more precise with respect to disciplinary conventions in the creative arts disciplines. These disciplines have been overlooked in most of the English for Specific Academic Purposes teaching materials.

Keywords

academic writing, lexical bundles, corpus analysis, genre, needs analysis

1 Introduction

Over the last decade, new PhD study programmes in visual arts and design have started to offer engaging opportunities to develop research skills that support professional practice in the creative industry. This phenomenon of research, as well as academic careers within the fields of arts, design and other creative industries such as the film industry, is now in focus because the creative industry offers more and more job positions, mostly in digital arts and design, industrial and product design, and 3D animation. Research in these disciplines attracts new students who have completed undergraduate studies but also those who return to university from well-established artistic careers. A simple query on a PhD Study portal (phdportal.com, 2024) shows over 450 PhD or doctoral programmes in visual arts in Europe ranging from media arts and creative technologies, film studies, materials research in industrial design to fashion design or photography. Most of the students in such research programmes are challenged to learn academic writing conventions in English, which leads them to international publishing in English language. Research typically carried out

Discourse and Interaction 18/2/2025, pp. 57-79
ISSN 1802-9930
https://doi.org/10.5817/DI2025-2-57

in these three disciplines is often interdisciplinary, placed on the edge of social and natural sciences. Furthermore, research issues within visual arts and design are rather new in academia and those who strive to publish within the fields of photography, product/industrial design, and audiovisual arts must look for academic conventions and characteristic features of relevant written discourse. It is the differences between writing norms in these three disciplines that this study aims to investigate.

The corpora compiled for the research, each comprising 30 published articles from photography, product/industrial design and audiovisual arts, were explored using a comparative method. Both quantitative and qualitative research approaches were used in this study. The investigated lexical bundles touch on matters of grammatical (lexico-grammatical), pragmatic (the functions of the bundles, coherence, accuracy of statements) and sociolinguistic (appropriateness of the choice of verbal means with regard to the situation) competence in writing. According to Bachman and Palmer's (1982) model of communicative competence, mastery of foreign language vocabulary is a component of all of these competencies. Research on the needs of doctoral students in academic writing with regard to the correct use of lexical bundles falls within the aforementioned areas. Findings from this corpus investigation can inform specialised academic writing courses and instructional materials that will support the enhancement of the quality of academic writing in English. As stated by Biber and Conrad (2009), research focused on description of the language used in registers and genres from a particular profession or an academic discipline is essential when preparing syllabi for English for Specific Academic Purposes (ESAP).

Multi-word sequences, also lexical bundles, have received a variety of labels and definitions, such as lexical phrases, formulas, pre-fabricated patterns, and n-grams. Most studies have the same focus on how words line up into more or less fixed combinations. Originally theoretical, from the 1990s most research of this type has been empirical (Gray & Biber, 2015, p. 125). Lexical bundles in the English language are not a linguistic exception; they are not just a peripheral domain, such as idioms. There are more than 40 labels used for the term multiple word bundles in English terminology (Wray, 2002; Wray & Perkins, 2000). Due to the focus on corpus analysis in the academic writing context, the working definition in this paper is adopted from Biber and Barbieri (2007). They simply define lexical bundles as "the most frequently recurring sequences of words" (2007, p. 264). These sequences are usually not complete in their structures and do not have an idiomatic meaning, but serve important discourse functions (Biber & Barbieri, 2007, p. 264).

It has been proven that multi-word sequences (*it is possible to, at the beginning of*) are important for the acquisition and production of both native and foreign languages (Christiansen & Arnon, 2017; Isbilen & Christiansen, 2018). Lexical bundles have an important role in both spoken and written registers; however, their effective acquisition and correct use are among the most difficult aspects of the foreign language learning process (Cortes, 2006; Paquot, 2010). When learning academic writing, students need to be made aware of lexical collocations and the occurrence of words in longer meaning units. This research adopts the Firthian (1957) theory, well phrased in the statement: "you shall know a word by the company it keeps" (p. 11). Cook (2021) points out that focusing on the complexity of lexical meaning is more important than looking for an isolated semantic category. It is not possible to isolate individual meanings of words; there is a general agreement that meaning is context dependent.

Previous research in ESAP argues for the need to focus on variation in discursive practice across disciplines (Cortes, 2004, 2006, 2023; Charles, 2015; Flowerdew, 2016; Hyland, 2008b, 2008c; Samraj, 2024). The main focus is on texts published in peer-reviewed journals, and the premise is that published articles follow disciplinary generic norms and generic attributes. As pointed out by Goulard et al. (2022), intra-generic variation concerns theoretical and empirical articles, position papers, case studies, literary reviews, critiques, design specifications, and more. Student written genres in academic contexts often have similar communicative purposes to published academic articles. Therefore, it is practical to look closer at the characteristics of published texts in the relevant disciplines which have their own specific writing conventions. Disciplinary generic norms can be characterised not only by describing the organisation of the text and its internal structure, but also by determining the rhetorical function and frequency of lexical bundles (Hyland, 2005, 2008b, 2008c, 2012; Hyland & Tse, 2009). In this paper, to characterise a published article in peer-reviewed journals in the three selected arts disciplines, the occurrence of four-word lexical bundles is identified, then their pragmatic functions are analysed in order to compare how authors in the three disciplines report on their research work. The aim is to identify lexical preferences in scholarly writing in the three target arts disciplines.

2 Lexical bundles in academic writing

Academic writing shares only a few lexical bundles with spoken academic discourse and even fewer with fiction and conversation. Such differences have been confirmed by studies of large corpora (Biber et al., 1999; Hyland, 2012; Hyland & Jiang, 2018; Simpson-Vlach & Ellis, 2010) mostly to identify how lexical bundles vary in frequency, form, and function by mode, discipline, and genre. As pointed out by Biber and Conrad (2009, p. 256), a different

communicative purpose of a text will use different sets of linguistic features that are functionally suitable for such a purpose. The research presented in this paper focuses specifically on lexical bundles as 4-grams which make up only 4% of spoken and approximately 2% of written academic discourse, yet they are among the useful building blocks of academic discourse (Biber et al., 2004). Multiple word bundles or lexical bundles are further defined by Biber and Barbieri (2007) as lexical clusters or bundles that occur in the form of a fixed sequence of words. They are usually not structurally complete and do not have idiomatic meaning (e.g., *the extent to which, at the same time, it should be noted*) but serve as important discursive devices in both spoken and written discourse (Biber & Barbieri, 2007). The aim of this study is to characterise the three selected arts disciplines in their professional and academic written discourse by examining the use of lexical bundles in articles published in peer-reviewed journals. It draws on previous studies by Wray (2002), Cortes (2004, 2006) and Hyland (2004, 2008a, 2008b), who confirmed that established formulae allow language users to express their identity as a particular group, the community of a particular discipline. The frequency of occurrence and rhetorical functions of lexical bundles depend on the discipline and its discourse conventions (Hyland, 2004, 2005, 2008a, 2008b, 2012; Hyland & Jiang, 2018), among other reasons, because research articles in different disciplines may have different communicative purposes associated with specific rhetorical conventions. In her research oriented to academic language and the genre of research papers in specific disciplines, Cortes (2004, 2006) points out that the use of established vocabulary signifies competent mastery of language within a particular register, or learning the conventions of a particular register also implies knowledge of particular established formulas. Previous published studies focusing on lexical bundles in academic discourse across disciplines (Chang, 2012; Hyland, 2002a, 2002b, 2004, 2005, 2008a, 2008b; Hyland & Jiang, 2018) support the notion underlying the research presented here: that the three disciplines likely vary to some extent in the lexical bundles they use.

3 Methodology

3.1 Research design

The mixed design approach applied in this corpus-based study encompasses quantitative and qualitative methods as described in Section 3.3 below. The three arts disciplines are analysed with the use of pre-selected taxonomies of lexical expressions in the register of academic writing within the genre of peer-reviewed journal articles. The two research questions for this corpus study aim at the realisation of lexical bundles that were retrieved as four-word units (i.e., 4-grams). These are first collected with the use of the Sketch Engine

software (Kilgarriff et al., 2014) and further identified as pragmatic markers that have discourse-functional properties. The research questions were formulated as follows:

1. What are the most frequent realisations of lexical bundles (limited to 4-grams) and their formal types in each of the three arts disciplines?
2. What rhetorical functions are presented in the most frequent lexical bundles (limited to 4-grams) in each of the three arts disciplines?

3.2 Composition of the corpora

The study analyses three corpora representing articles published in peer-review journals in the three disciplines. To limit the study, only authors from universities in English-speaking countries were selected. They have been categorised as expert users of English by name and affiliation with an institution in the UK, Canada, Australia or the USA. These authors have published research in English and are researchers, scholars or art theorists with specialisations in photography, product/industrial design, aesthetics, philosophy, digital media theory and visual communication. The selection was made from ten journals, three articles from each journal. The composition is described in Table 1. The total number of tokens is highest in the discipline of photography, as these are longer essays, with a disputative, descriptive and argumentative purpose. Yet a slightly higher number of unique word items and lemmas is observed in the discipline of audiovisual arts. This is in contrast to the texts published in product/industrial design, not only because published articles in this discipline are on average shorter. Also, when comparing the number of unique word items, it can be concluded that the articles in the disciplines of photography and audiovisual arts are lexically more varied than articles in product/industrial design.

Corpus type	Name	Genre	Number of texts	Size
L1 speakers	Normative Corpus 1	Published articles in photography	30	**243,380 words** 293,044 tokens 23,758 unique word items 17,089 lemmas
L1 speakers	Normative Corpus 2	Published articles in product/ industrial design	30	**180,218 words** 218,049 tokens 16,320 unique word items 11,709 lemmas
L1 speakers	Normative Corpus 3	Published articles in audiovisual arts	30	**222,394 words** 265,588 tokens 23,985 unique word items 17,109 lemmas

Table 1: Corpus characteristics and size comparison

The journals were selected as a representative sample of texts that meet the reading and writing needs for PhD students in the selected disciplines.

For Corpus 1 the articles take a historical perspective, refer to theory of photography as well as photographic artistic practice, the topics discussed in the articles relate to digital as well as analogue photography and their theoretical and practical problematics (*History of Photography, Photographies, Photography and Culture, Journal of Photography Education, Journal of Aesthetics and Art Criticism, Oxford Art Journal, Critical Inquiry, Parallax, Arts, American Art,* and *Representations)*.

Corpus 2 is composed of papers in the disciplines of product/industrial design and related technical fields such as industrial engineering, from which product designers draw their information (*Design Issues, Co-Design: International Journal of Co-Creation in Design and the Arts, International Journal of Design, The Design Journal, Human-Computer Interaction, International Journal of Sustainable Engineering, Journal of Industrial Ecology, Journal of Engineering Design, International Journal of Art and Design Education,* and *Journal of Foot and Ankle Research*). The topics vary from product life cycle, sustainability and materials ecology in product design, industrial manufacturing and digitisation, designer interaction with the digital environment, designer education and studio work, collaborative design and the product design process, to women as designers.

The last discipline, the audiovisual arts as Corpus 3, covers a rather wide range of topics from film theory and history to the problematics of film production process, screenwriting, roles of individual crew members and technical processing, digital animation, VFX effects processing and more (*Journal of Film and Video, Media Practice and Education, Animation Studies, Journal of Cinema and Media Studies, Cinema Journal, Film Criticism, Film History, Literature and Film Quarterly, Film Matters,* and *Journal of Screenwriting*).

Many art journals specialise in selected topics for a longer period; their issues are often monothematic. To compile complex and representative corpora 1, 2 and 3 that would correspond to the varied topics of the three disciplines, the selected journal articles come from a longer period of publication, a period that spans from 2009 to 2023. A shorter period of publication would not allow a wide enough selection of topics and could be over-represented by the most recent themes, such as gender studies, impact of Covid-19, or research carried out with utilisation of recent digitalisation and programming. The starting point of 2009 is when digitalised arts and design journals began to cover a spectrum of themes that broadly correspond to the research topics in the target group of PhD students. Another selection criterion which was considered when building the corpus was

whether the authors are expert users of English. A positive assumption was made after checking their home institution affiliation, previous publishing activity, and their names. After the texts were selected, they were stripped of figure captions, photographs, tables, bibliographic references and references to other sources before the corpus file was closed. Then, with the use of the Sketch Engine software package (Kilgarriff et al., 2014), the three corpora were compiled.

3.3 Procedure

The method of quantitative and qualitative corpus analysis was chosen to determine the frequency of occurrence of lexical bundles forms. The four-word bundles were chosen for their length compared to shorter phrases, which are admittedly more numerous in occurrence; three-word bundles occur up to ten times more frequently in spoken and written discourse according to Biber (2009), but four-word phrases are more applicable for the analysis of rhetorical functions. Only forms with a minimum occurrence of 20 per million words were retrieved from the data sets. The Sketch Engine software package (Kilgarriff et al., 2014) was used for the analysis with the use of the implemented software statistical tools. The applied calculations for Log Likelihood (critical value 3.84) and the effect size statistics Log Ratio and %DIFF (Rayson & Garside, 2000; McEnery & Hardie, 2024) were done with the online calculator available from Lancaster University web pages (https://ucrel.lancs.ac.uk/llwizard.html). The data calculations and the procedure followed the guidance which was devised by Rayson (2024) and Centre for Corpus Approaches to Social Sciences at Lancaster University (https://cass.lancs.ac.uk/log-ratio-an-informal-introduction/). The occurrence of 4-grams together with the concordance lines of each 4-gram were examined. When reading the concordances, the context of 4-gram occurrence was monitored to eliminate content-specific nominal phrases that referred to artwork and film titles, journal titles, art associations, and so forth (e.g., *national film board of, This Is the American Earth, 'Photography and the New God'*). As highlighted in a previous survey by Davis and Morley (2015, p. 24), longer lexical bundles, those above 5 words, were mostly found unsuitable for re-use in academic writing as they contain non-generic content words that are tied to a particular topic or idea. Thus, lexical bundles composed of generic content words found in the Academic Word List or the General Service List were selected for the formal taxonomy analysis.

For functional analysis, only four-word bundles that meet Hyland's (2005, 2008a) taxonomy for rhetorical functions were marked to analyse the rhetorical functions of the most frequent 4-grams. To ensure representativeness of the corpus data, the even and uneven distribution of 4-grams across the corpus

in the individual texts was also observed, and instances of high reoccurrence with a standard deviation of 3 and more within a single text were recognised, isolated and ignored. The applied corpus study restrictions assure comparability with previous research. Obviously, the focus on 4-grams has restricted the scope of lexical bundles under analysis, since as Conrad and Biber (2005, p. 69) concluded, many text-organising bundles and topic elaborations are realised by 2- and 3-grams (*for example*, *such as*, *in order to*, *in terms of*). The size of the three individual corpora sets is compared at several levels, with tools that are available in the Sketch Engine package (Kilgariff et al., 2014). Comparisons were made according to the number of lemmas, the representative lexical forms, the range of word forms (number of unique textual forms) and tokens (absolute occurrence of these unique textual forms). Normalisation is in a standard format, per million words, referred to forthwith as wpm.

3.4 Taxonomies for lexical bundles

As mentioned above, Biber et al. (1999) distinguish between lexical bundles for spoken and written registers with separate taxonomies for each concluding that bundles in academic prose are phrasal rather than clausal. Biber et al. (2003) classify written academic bundles into the categories of *noun phrase fragments*, *prepositional phrase fragments and other prepositional phrase fragments*, *anticipatory It + verb/adjective phrase*, *verb/adjective + that/to clause fragment and passive verb + to clause*, *passive verb + prepositional phrase*. The formal taxonomy by Cortes (2004) for analysing lexical bundles in the disciplines of biology and history has eight slightly different categories. For example, Cortes identifies *noun phrases with "of" phrase fragment* as a single category (*a wide range of*), different from *prepositional phrase with embedded "of" phrase* (*as a consequence of*).

In this study the following categories were drawn:

1. nominal and prepositional phrase fragments + *of*, *in*, *at* (*of the design process*, *in the context of, at the same time*),
2. other expressions, prepositional phrases and phrase fragments (*as well as the*),
3. passive verb + *to* clause, passive verb + prepositional phrase (*is reported to be, is based on the),*
4. modal verbs + *be* + participle (*can be found in*),
5. the use of the pronoun *it* as an anticipatory *it* structure (*it is important to*).

Conrad and Biber (2005, p. 58–59) point out that lexical bundles are not a by-product of corpus-frequency analysis, and moreover, they are consistently functional. They serve basic discourse functions related to the expression of

stance (*it is necessary to*, *the fact that the*), discourse organisation (*on the other hand*) and referential framing (*as shown in Figure 4*, *in the case of*), and show differences from other lexico-grammatical features in their patterns of use as they are influenced by both mode and communicative purpose (Biber & Barbieri, 2007, p. 265–282). The rhetorical functions of lexical bundles are examined in the pragmatic approach to language as a means of social interaction, which consists of a negotiation between the author of a text and its reader. Research published so far in this area has suggested several categories for the rhetorical functions of lexical bundles. This study adopts Hyland's (2005, 2008a, 2008b, 2023) taxonomy of lexical bundles, which draws on Halliday's ideational, interpersonal and textual functions of language. Hyland's taxonomy of rhetorical functions for lexical bundles distinguishes between:

1. research-oriented (*in the present study*),
2. text-oriented (*on the other hand*),
3. participant-oriented (*it is possible that*).

The functional-pragmatic qualitative analysis in this study focuses on the linguistic means of expressing the above stated functions. Hyland (2008) further divided the interpersonal function, that is, the participant-oriented rhetorical function, into two sub-categories:

1) Stance features: stance features characterise lexical devices used to express the author's attitude and evaluation (*are likely to be*, *it is likely to*, *it would certainly appear*, *may be due to,* etc.); these are expressions of what the author thinks or assumes (*assumes to be*, *it will be*, *it was the case*). They can also include devices that mitigate the author's claims (hedges) and devices that enhance the credibility of the author's statements (boosters).
2) Engagement features: devices for engaging the reader that directly address the reader. The author perceives that there is something to point out, something to highlight (*it is worth pointing out*, *it is noteworthy*, *it is important to*, *it is necessary to*).

Lexical bundles for these two sub-categories are acknowledged in this study under the participant-oriented function.

4 Findings

4.1 Formal typology of lexical bundles

A comparison of the formal types of lexical bundles for the three disciplinary corpora, the first three examples having the highest frequency of occurrence, are shown in Tables 2, 3 and 4 below. The overall frequencies are presented together with numbers for identified types of lexical bundles in each observed category.

The second column also shows the absolute frequency of occurrence, and the third column gives normalised frequency per million words. The threshold frequency is 20 per million words.

Corpus 1 Photography	**Raw instances**	**Normalised data**
Nominal and prepositional phrase + *of*, *in*, *at*	**Total 67 form types, 651 occurrences**	**2,221.50**
at the same time	33	112.61
the way in which	25	85.31
in the case of	21	71.66
Other expressions, prepositional phrases and phrase fragments	**Total 11 form types, 132 occurrences**	**450.44**
on the other hand	30	102.37
on the one hand	23	78.49
as well as the	18	61.42
Passive verb + *to* clause, passive verb + prepositional phrase	none	0
Modal verb + *be*+ 3rd verb form	**Total 1 form type, 8 occurrences**	**27.29**
can be found in	8	27.30
can be understood as	5	17.06
can be seen in	5	17.06
Anticipatory *It*	**Total 5 form types, 32 occurrences**	109.19
it cannot be + (verb form)	8	27.30
it (be) important to	6	20.47
this is not to (say, suggest, deny)	6	20.47

Table 2: Corpus 1. Formal types of 4-grams. Raw occurrences and normalised relative frequencies

Corpus 2 **Product/industrial design**	**Raw instances**	**Normalised data**
Nominal and prepositional phrase + *of*, *in*, *at*	**Total** **68 form types,** **477occurrences**	2,187.58
of the design process	17	77.96
in the context of	16	73.83
in terms of the	14	64.21
Other expressions, prepositional phrases and phrase fragments	**Total** **10 form types,** **104 occurrences**	**476.95**
as well as the	30	137.58
on the other hand	16	73.38
the extent to which	13	59.62
Passive verb + *to* clause, passive verb + prepositional phrase	**Total** **1 form type,** **6 occurrences**	**27.51**
is based on the	6	27.52
Modal verb + *be* + 3rd verb form	**Total** **4 form types,** **30 occurrences**	**137.58**
can/should be used to	11	50.45
to be able to + verb	7	31.10
can be found in	6	27.52
Anticipatory *It*	**Total** **7 form types,** **45 occurrences**	**206.37**
it (be) important to	10	45.86
it is possible to	7	31.10
it should be noted (that)	6	27.52

Table 3: Corpus 2. Formal types of 4-grams. Raw occurrences and normalised relative frequencies

Corpus 3 Audiovisual arts	Raw instances	Normalised data
Nominal and prepositional phrase + *of*, *in*, *at*	**Total 48 form types, 457 occurrences**	**1,720.71**
at the same time	32	120.41
the end of the	24	90.37
in front of the	24	90.37
Other expressions, prepositional phrases and phrase fragments	**Total 12 form types, 98 occurrences**	**368.99**
as well as the	17	64.01
for the first time	10	37.65
the extent to which	10	37.65
Passive verb + *to* clause, passive verb + prepositional phrase	**Total 1 form type, 6 occurrences**	**22.59**
be seen as a	6	22.59
Modal verb + *be* + 3rd verb form	**Total 3 form types, 18 occurrences**	**67.77**
can/could be seen as	6	22.59
one could argue that	6	22.59
that we cannot (can not)	6	22.59
Anticipatory *It*	**Total 1 form types, 8 occurrences**	**30.12**
It is clear that	8	30.12

Table 4: Corpus 3. Formal types of 4-grams. Raw occurrences and normalised relative frequencies

The data presented in Tables 2, 3 and 4 indicate that the disciplines of photography and product/industrial design use nominal and prepositional phrases in high frequencies (2,221.50 and 2,187.58 respectively). However, in audiovisual arts studies the occurrence of these phrases is less frequent (1,720.71). Modals, typically associated with hedging and boosting expressions, are the most frequent in product/industrial design (137.58), followed by audiovisual studies, where they are half as frequent (67.77). They are least frequent in photography (27.29). The 4-grams containing the anticipatory *it* structure are again most frequent in product/industrial design (206.37), followed by photography (109.19) and audiovisual arts (30.12). This shows that texts in the discipline of product/industrial design use formulaic language the most.

The comparative corpus research done in 2018 by Hyland and Jiang identified that the number of bundles in the disciplines of applied linguistics, engineering, sociology and biology did not increase over the observed time and the most frequent among the top 20 bundles in academic writing over the past 50 years were *on the other hand, on the basis of, in the case of, at the same time, in terms of the, in the presence of, in the form of* and *in the absence of* (Hyland & Jiang, 2018). The bundle *on the other hand* is also at the top for photography and product/industrial design disciplines in this study. However, in audiovisual arts this bundle did not reach the three top positions.

The use of verbs in passive voice in lexical bundles did not meet the threshold frequency of 20 per million words and the limitations that were set to 4-grams. This could lead to a biased view on the usage of expressions in passive voice. To eliminate such bias and to look closer at this language feature, Figure 1 illustrates the number of verbs in passive voice across the three disciplines. The normalised data (Y-axis) indicate that the passive, impersonal constructs, are mostly typical for the discipline of product/industrial design (27,181) and occur in similar frequencies in photography (21,068) and audiovisual arts (19,906). A closer examination of the verb types used in passive voice also revealed that their frequency does not necessarily have a direct correlation with their occurrence in 4-grams.

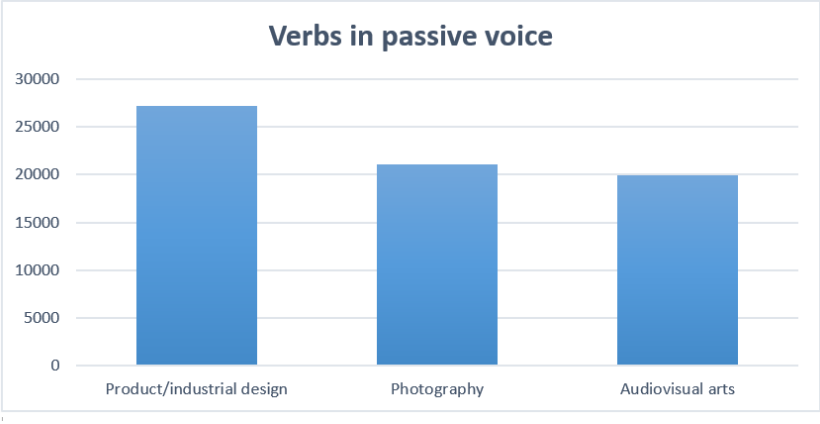

Figure 1: Verbs in passive voice, disciplinary differences (normalised data)

Table 5 shows a comparative analysis of selected formal types, based on Conrad and Biber's (2005) taxonomy for academic prose, for the realised 4-grams

(threshold 20 per million words) in three corpora with statistical indicators. The three corpora are compared, always with a larger set to a smaller set of tokens: *photography* to *product/industrial design*, *audiovisual arts* to *product/industrial design*, *photography* to *audiovisual arts*. The smallest set, the discipline of product/industrial design, shows the larger relative frequency of the 4-grams in most instances.

	Photo-graphy	Product/ industrial design	Audio visual arts	Product/ industrial design	Photo-graphy	Audio visual arts
Nominal and prep phrases + *of*, *in*, *at*						
Raw instances	651	477	457	477	651	457
Log Likelihood		**0.03**	**14.94**		**18.90**	
Log Ratio		**0.02**	**-0.37**		**0.38**	
%DIFF		**1.06**	**-22.36**		**30.17**	
Other prepositional phrases						
Raw instances	132	104	98	104	132	98
Log Likelihood		**0.22**	**3.67**		**2.45**	
Log Ratio		**-0.09**	**-0.39**		**0.30**	
%DIFF		**-6.02**	**-23.64**		**23.08**	
Verb in passive						
Raw instances	0	6	6	6	0	6
Log Likelihood	46.15		**0.13**		**8.87**	
Log Ratio	**-6.19**		**-0.30**		**-3.72**	
%DIFF	**-100**		**-18.96**		**-100**	
Modal verb + *be* + 3rd verb form						
Raw instances	8	30	18	30	8	18
Log Likelihood		**21.03**	**6.09**		**4.90**	
Log Ratio		**-2.34**	**-1.04**		**-1.30**	
%DIFF		**-80.25**	**-51.38**		**-59.39**	
Anticipatory It						
Raw instances	32	45	8	45	32	8
Log Likelihood		**7.84**	**36.86**		**13.34**	
Log Ratio		**-0.93**	**-2.80**		**1.87**	
%DIFF		**-3.02**	**-85.59**		**265.51**	

Table 5: Formal types of 4-grams, raw instances. Log Ratio provided for effect-size, with a minus sign indicating difference in underuse for the first corpus in the pair

What stands out from the data in Table 5 is that the texts in the discipline of product/industrial design contain more lexical bundles than the other two disciplines, photography and audio-visual arts. The closer numbers of realisations for the nominal and prepositional phrases with *of, in* and *at* for the corpora of photography and product/industrial design indicate similar usage of formalised prepositional phrases, a typical characterisation of formal academic discourse in the hard science and technology fields. These findings confirm previous studies (Hyland, 2012, 2023) which indicate that texts in technical and engineering fields use prefabricated structures more often than texts in humanities fields. The question is, where to fit the discipline of photography? Its technical and technological expertise as much as a philosophical background in theoretical studies reflect knowledge from both technical and humanities fields. Texts from the discipline of photography prove to be interdisciplinary and themes discussed in photography journals stretch across both categories.

The most frequent type of noun phrase and prepositional phrase fragments is in photography, but the difference in the frequency of this type of bundle between photography and product/industrial design is not statistically significant. Significant differences are observed, such as a lower frequency of bundles with modal verb + *be* + 3rd verb form in photography (-80.25 %DIFF) and audiovisual arts (-51.38 %DIFF), in comparison to product/industrial design. This indicates stronger utilisation of boosting and hedging language in the product/industrial design texts, which fits with the so far researched typical text characteristics of technical and hard science disciplines. A lower frequency of the anticipatory *it* structure for audiovisual arts (-85.59 %DIFF) is also seen in comparison to product/industrial design. The distancing function of the anticipatory *it* structure or *it-clause* in academic writing is mostly used in a rhetorical function which leads to an impression of the presentation of objective, impersonal knowledge (Hewings & Hewings, 2002). Researchers' claims and standpoints position them more firmly in the disciplinary discourse which is more individualistic. As Hyland (2023, p. 87) suggests, writers in the social sciences and humanities "put themselves into the text more" when compared with hard science disciplinary discourses. The findings also point to a statistically significant overuse of the anticipatory *it* structure (265.51 %DIFF), nominal and prepositional phrases with *of, in, at* (30.17 %DIFF) and other prepositional phrases (23.08 %DIFF) for photography in comparison to audiovisual arts. This can be attributed to a more conversational style of writing in the audiovisual arts.

4.2 Functional typology of lexical bundles

The rhetorical functions of lexical bundles are examined in the pragmatic approach to language as means of social interaction, which consists of a negotiation between the author of a text and its reader. Data presented in Tables 6, 7 and 8 indicate functional type occurrence of 4-gram lexical bundles in the three disciplines.

Corpus 1 Photography	**Raw occurrences**	**Normalised data**
Content-oriented	**Total 65 form types, 791 occurrences**	**2,699.25**
at the same time	33	112.61
the way in which	25	85.31
in the history of	25	85.31
Text-oriented	**Total 12 form types, 147 occurrences**	**501.63**
on the other hand	30	102.37
on the one hand	23	78.49
in the case of	21	71.66
Participant-oriented	**Total 6 forms, 47 occurrences**	**160.38**
can be found in	8	27.30
it cannot be (ignored, denied, understood, etc.)	8	27.30
appears to be a (basic feature, fortunate side)	7	23.89

Table 6: Functional types, most frequent examples in raw occurrences and relative frequencies

Corpus 2 Product/industrial design	**Raw occurrences**	**Normalised data**
Content-oriented	**Total 42 form types, 309 raw occurrences**	**1,417.11**
of the design process	17	77.96
in the development of	11	50.45
the development of a	11	50.45
Text-oriented	**Total 17 form types, 131 raw occurrences**	**600.78**
on the other hand	16	73.38
in terms of the	14	64.21
in the case of	13	59.62

Corpus 2 Product/industrial design	Raw occurrences	Normalised data
Participant-oriented	**Total 17 form types, 113 occurrences**	**518.23**
it is important to	10	45.86
can be used to	9	41.28
there is a need	8	36.69

Table 7: Functional types, most frequent examples in raw occurrences and relative frequencies

Corpus 3 Audiovisual arts	Raw occurrences	Normalised data
Content-oriented	**Total 50 form types, 470 occurrences**	**1,769.65**
at the same time	32	120.41
the end of the	24	90.37
in front of the	24	90.37
Text-oriented	**Total 14 form types, 116 occurrences**	**436.76**
in the case of	20	75.30
as well as the	17	64.01
on the one hand	11	41.42
Participant-oriented	**Total 5 form types, 33 occurrences**	**124.25**
this is not to (suggest, imply, say)	8	30.12
not to suggest that	7	26.36
one could argue that	6	22.59

Table 8: Functional types, most frequent examples in raw occurrences and relative frequencies

Collected data show that content-oriented phrases were used the most in photography (2,699.25) and the least in product/industrial design (1,417.11). Text-oriented bundles have the highest occurrence in product/industrial design (600.78) and the lowest in audiovisual arts (436.76). Participant-oriented bundles were used most in product/industrial design (518.23), and least in audiovisual arts (124.25). Comparing and sorting the retrieved 4-grams on concordance lines with the functional taxonomy by Hyland (2008a, 2008c, 2012) who composed his corpora from texts in disciplines other than photography, product/industrial design and audiovisual studies, revealed several differences in their utilisation. For example, the bundle *can be seen in* when followed with *Figure 1, in the table, data* etc. would fall into the category of text-oriented framing signals in many scientific disciplines, but in the three corpora here it falls into the content-oriented category: *can be seen in* would be followed for example with

a debate, a description, a late 13th century, Avatar, its entirety on YouTube. Most framing signals refer to *concept of, nature of, presence of...*, while description and content-oriented bundles often refer to *forms, ways, relationships, figure(s).*

Findings in the functional comparative analysis are illustrated here in Table 9.

	Photo-graphy	**Product/ industrial design**	**Audio visual arts**	**Product/ industrial design**	**Photo-graphy**	**Audio visual arts**
Content-oriented						
Raw instances	65	42	50	42	65	50
Log Likelihood	**0.48**		**0.03**		**0.84**	
Log Ratio	**0.20**		**-0.05**		**0.25**	
%DIFF	**14.60**		**-3.53**		**18.79**	
Text- instances						
Raw instances	12	17	14	17	12	14
Log Likelihood	**3.02**		**1.26**		**0.39**	
Log Ratio	**-0.94**		**-0.58**		**-0.35**	
%DIFF	**-47.73**		**-33.26**		**-21.68**	
Participant-oriented						
Raw instances	6	17	5	17	6	5
Log Likelihood	**9.30**		**9.68**		**0.02**	
Log Ratio	**-1.94**		**-2.07**		**0.13**	
%DIFF	**-73.87**		**-76.17**		**9.65**	

Table 9: Functional taxonomy, raw occurrences. Log Ratio provided for effect-size, with a minus sign indicating difference in underuse for the first corpus in the pair

To conclude the discussion of the functional types across the three disciplines, it should be noted that the utilisation of the rhetorical functions shows differences that correspond with previous findings in the formal analysis.

Participant-oriented lexical bundles in 4-grams, here mostly represented with the use of the pronoun *it* as an anticipatory *it* structure, show the most statistically significant difference when comparing product/industrial design with the two remaining disciplines. It is more significantly represented in the corpus from product/industrial design (504 wpm) compared to the corpus from photography (130 wpm) and audiovisual arts (122 wpm). A stronger distancing orientation away from the participants in hard sciences with the anticipatory *it* structure was also observed in a previous study by Hyland and Jiang (2018), who point out that over the observed period of time the hard sciences reduced their use of bundles which focus on reporting research and started adopting more forms which carry

epistemic evaluations, attitudes or modal meanings. This study confirmed more distancing lexical bundles with the use of participant-oriented functions for product/industrial design. Furthermore, closer reading of the concordance lines for individual occurrences within texts revealed a surprisingly high frequency, even an overuse, of lexical bundles by some authors in the product/industrial design corpus. Essay-type articles in photography and research papers in product/industrial design discipline have similar communicative purposes to persuade, speculate, or argue, but their authors use different rhetorical strategies.

The above collected data can be compared with previous research, specifically that language means for distancing can vary with respect to scientific discipline (Hyland, 2004) as can rhetoric used for argumentation (Walková & Bradford, 2022), also with respect to grammatico-syntactic level of language in which the text is written (Dontcheva-Navratilova, 2013a; Jančaříková, 2023). The differences can be also significant due to different cultural and academic conventions of the relevant discourse and readiness to adapt to certain academic conventions. For example, novice writers whose L1 is Czech or German differ in the way they adapt to Anglo-American writing conventions (Dontcheva-Navratilova, 2013b, 2023). Such findings contribute to discussions in ESAP lessons aimed at academic writing and individual voice in scientific disciplines.

5 Conclusion

The aim of this study was to explore how the occurrence of lexical bundles contributes to the characterisation of texts produced in different disciplines. The findings build on previous studies in academic written English (Biber et al., 2004; Cortes, 2004, 2006; Conrad & Biber, 2005) and specifically those focusing on the differences in the language used in written academic discourse across disciplines (Hyland, 2005, 2008a, 2008b, 2008c, 2023; Hyland & Jiang, 2018).

With reference to the research questions, the most frequent realisation of lexical bundles (4-grams) in their formal types in all three disciplines are nominal and prepositional phrases with *of, in, at*. It has been demonstrated that authors of texts in photography and product/industrial design express themselves more frequently with the use of lexical bundles which they use in a wider variety than the authors in audiovisual arts. The findings have also pointed to a statistically significant overuse of nominal phrases and prepositional phrases *of, in, at* for photography in comparison to audiovisual arts. This can be attributed to a more open or conversational style of writing in the audiovisual arts. Technical and technological expertise as much as a philosophical background of theoretical studies in photography reflect knowledge from both technical and humanities fields. Photography texts prove to be interdisciplinary, with themes discussed in

photography journals spanning both categories. In contrast, greater utilisation of hedging or boosting language and distancing in the product/industrial design texts, which aligns with the so far researched typical text characteristics of technical and hard science disciplines, shows that product and industrial designers have a closer affiliation with technical fields. The tendency for these authors to use the lexical bundles of the anticipatory *it* structure more frequently suggests that in this discipline authors maintain a greater distance from the content of the text in order to create an impression of objectivity. Findings in the functional taxonomy also pointed out that the use of modals, indicators of boosting and hedging, is less frequent in photography and the audiovisual arts than in product/industrial design. As mentioned by Hyland and Jiang (2018), authors in the soft sciences tend to make stronger claims. Photography and audiovisual arts disciplines often opt for creative writing styles, with less utilisation of phraseology that is found in ESAP teaching materials for hard science students.

The study results can be utilised in the preparation of teaching materials for ESAP classes where the academic written discourse of arts and design disciplines is relevant for the course participants. Discipline-specific instructional materials can help students understand how authors guide readers through the text, what stance-taking strategies they adopt, or how they report on research data. Despite the wide range of resources currently available via writing manuals and phrasebanks or in the form of generative AI tools that supply novice writers in English with immediate support, the results of discipline-specific studies have the benefit of being tailor-made for specific PhD students' writing needs. Students can become more aware of the differences in academic genre characteristics and enhance their academic writing skills, rather than being dependant on a blind re-use of those skeletal phrases that can be readily added with generative AI tools today. Tailored study materials should support a creative approach to writing which leads to individual writing styles, a highly valued skill in any genre.

References

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, *16*(4), 449–465. https://doi.org/10.2307/3586464

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*, 263–286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press. https://doi.org/10.1017/CBO9780511814358

Biber, D., Conrad, S., & Cortes, V. (2003). Towards a taxonomy of lexical bundles in speech and writing. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the Lune: A festschrift for Geoffrey Leech* (pp. 71–92). Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education. https://doi.org/10.1162/089120101300346831

CASS, ESRC Centre for Corpus Approaches to Social Science. (n.d.). *Log ratio—An informal introduction*. https://cass.lancs.ac.uk/log-ratio-an-informal-introduction/

Chang, P. (2012). Using a stance corpus to learn about effective authorial stance-taking: A textlinguistic approach. *ReCALL*, *24*(2), 209–236. https://doi.org/10.1017/S0958344012000079

Charles, M. (2015). Genre, corpus and discourse: Enriching EAP pedagogy. In P. Thompson & G. Diani (Eds.), *English for academic purposes: Approaches and implications* (pp. 265–284). Cambridge Scholars Publishing.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*(2), 542–551. https://doi.org/10.1111/tops.12274

Conrad, S. M., & Biber, D. (2005). Frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, *20*, 56–71. https://doi.org/10.1515/9783484604674.56

Cook, V. (2021). The language in language and thinking. *Vigo International Journal of Applied Linguistics*, *18*, 35–58. https://doi.org/10.35869/vial.v0i18

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391–406. https://doi.org/10.1016/j.linged.2007.02.001

Cortes, V. (2023). Lexical bundles and phrase frames. In M. Kessler & C. Polio (Eds.), *Conducting genre-based research in applied linguistics* (pp. 105–125). Routledge. https://doi.org/10.4324/9781003300847

Davis, M., & Morley, J. (2015). Phrasal intertextuality: The responses of academics from different disciplines to students' re-use of phrases. *Journal of Second Language Writing*, *28*, 20–35. https://doi.org/10.1016/j.jslw.2015.02.004

Dontcheva-Navratilova, O. (2013a). Grammar and discourse. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2324–2331). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0464

Dontcheva-Navratilova, O. (2013b). Lexical bundles indicating authorial presence: A cross-cultural analysis of novice Czech and German writers' academic discourse. *Discourse and Interaction*, *6*(1), 7–24. https://doi.org/10.5817/DI2013-1-7

Dontcheva-Navratilova, O. (2023). Self-mention in L2 (Czech) learner academic discourse: Realisations, functions and distribution across master's theses. *Journal of English for Academic Purposes*, *64*, 1–14. https://doi.org/10.1016/j.jeap.2023.101272

Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford University Press.

Flowerdew, J. (2016). English for specific academic purposes (ESAP) writing. *Writing Pedagogy*, *8*(1), 1–4. https://doi.org/10.1558/wap.v8i1.30051
Goulard, L., Biber, D., & Reppen, R. (2022). In this essay, I will…: Examining variation of communicative purpose in student written genres. *Journal of English for Academic Purposes*, *59*(3), 13–38. https://doi.org/10.1016/j.jeap.2022.101159
Gray, B., & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 125–145). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377
Hewings, M., & Hewings, A. (2002). "It is interesting to note that…": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, *21*(4), 367–383. https://doi.org/10.1016/S0889-4906(01)00016-3
Hyland, K. (2002a). Directives: Argument and engagement in academic writing. *Applied Linguistics*, *23*(2), 215–239. https://doi.org/10.1093/applin/23.2.215
Hyland, K. (2002b). Specificity revisited: How far should we go now? *English for Specific Purposes*, *21*(4), 385–395. https://doi.org/10.1016/S0889-4906(01)00028-X
Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press ELT. https://doi.org/10.3998/mpub.6719
Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–191. https://doi.org/10.1177/1461445605050365
Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.x
Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. https://doi.org/10.1016/J.ESP.2007.06.001
Hyland, K. (2008c). Genre and academic writing in the disciplines. *Language Teaching*, *41*(4), 543–562. https://doi.org/10.1017/S0261444808005235
Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150–169. https://doi.org/10.1017/S0267190512000037
Hyland, K. (2023). Changing times; changing texts. In P. M. Rogers, D. R. Russell, P. Carlino & J. M. Marine (Eds.), *Writing as a human activity: Implications and applications of the work of Charles Bazerman* (Vol. 1, pp. 47–66). WAC Clearinghouse. https://doi.org/10.37514/PER-B.2023.1800.2.3
Hyland, K., & Jiang, F. K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, *23*(4), 383–407. https://doi.org/10.1075/ijcl.17080.hyl
Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies*, *9*(2), 111–129. https://doi.org/10.6018/ijes.9.2.90781
Isbilen, E. S., & Christiansen, M. H. (2018). Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science*, *12*(2), 713–726. https://doi.org/10.1111/tops.12376
Jančaříková, R. (2023). Attitude markers in L2 learners' academic writing: A case study of master's theses by Czech students compared to L1 students' writings. *Brno Studies in English*, *1*, 5–31. https://doi.org/10.5817/bse2023-1-1
Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, P. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9
McEnery, T., & Hardie, A. (2024, May 10). *Support website for Corpus Linguistics: Method, theory and practice*. http://corpora.lancs.ac.uk/clmtp

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Continuum.

Rayson, P. (2024, May 4). *Testing for significance: Log likelihood*. https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/l08_4.htm

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora, Held in Conjunction With the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000). 1–8 October 2000, Hong Kong* (pp. 1–6).

Samraj, B. (2024). Disciplinary differences in lexical bundles use: A cautionary tale from methodological variations. *Journal of English for Academic Purposes*, *70*, Article 101399. https://doi.org/10.1016/j.jeap.2024.101399

Simpson-Vlach, R. C., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Walková, M., & Bradford, J. (2022). Constructing an argument in academic writing across disciplines. *ESP Today*, *10*(1), 22–42. https://doi.org/10.18485/esptoday.2022.10.1.2

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1–28. https://doi.org/10.1016/S0271-5309(99)00015-4

Hana Atcheson is a lecturer at the Tomas Bata University in Zlín, Czech Republic. She is a coordinator and curriculum planner of English courses at the Faculty of Multimedia Communication. Her research interests are in pragmatic and lexical approaches to language teaching, corpus linguistics methods for TESOL, and genre analysis for ESAP.

ORCID ID: https://orcid.org/0000-0002-1221-989X

Address: Hana Atcheson, Language Centre, Faculty of Humanities, Tomas Bata University, Štefánikova 5670, Zlín 760 01, Czech Republic. [e-mail: atcheson@utb.cz]