

# ANGLICA

An International Journal of English Studies

30/2 2021

## EDITORS

Marzena Sokołowska-Paryż [m.a.sokolowska-paryz@uw.edu.pl]  
Anna Wojtyś [a.wojtys@uw.edu.pl]

## ASSOCIATE EDITORS

Silvia Bruti [silvia.bruti@unipi.it]  
Lourdes López Ropero [lourdes.lopez@ua.es]  
Martin Löschnigg [martin.loeschnigg@uni-graz.at]  
Jerzy Nykiel [jerzy.nykiel@uib.no]

## ASSISTANT EDITORS

Magdalena Kizeweter [m.kizeweter@uw.edu.pl]  
Dominika Lewandowska-Rodak [dominika.lewandowska@o2.pl]  
Bartosz Lutostański [b.lutostanski@uw.edu.pl]  
Przemysław Uściński [przemek.u@hotmail.com]

## ENGLISH LANGUAGE EDITOR

Barry Keane [bkeane@uw.edu.pl]

## ADVISORY BOARD

Michael Bilynsky, University of Lviv  
Andrzej Bogusławski, University of Warsaw  
Mirosława Bucholtz, Nicolaus Copernicus University, Toruń  
Jan Čermák, Charles University, Prague  
Edwin Duncan, Towson University  
Jacek Fabiszak, Adam Mickiewicz University, Poznań  
Elżbieta Foeller-Pituch, Northwestern University, Evanston-Chicago  
Piotr Gąsiorowski, Adam Mickiewicz University, Poznań  
Keith Hanley, Lancaster University  
Andrea Herrera, University of Colorado, Colorado Springs  
Christopher Knight, University of Montana,  
Marcin Krygier, Adam Mickiewicz University, Poznań  
Krystyna Kujawińska-Courtney, University of Łódź  
Brian Lowrey, Université de Picardie Jules Verne, Amiens  
Zbigniew Mazur, Maria Curie-Skłodowska University, Lublin  
Rafał Molencki, University of Silesia, Sosnowiec  
John G. Newman, University of Texas Rio Grande Valley  
Jerzy Rubach, University of Iowa  
Piotr Ruszkiewicz, Pedagogical University, Cracow  
Hans Sauer, University of Munich  
Krystyna Stamirowska, Jagiellonian University, Cracow  
Merja Stenroos, University of Stavanger  
Jeremy Tambling, University of Manchester  
Peter de Voogd, University of Utrecht  
Anna Walczuk, Jagiellonian University, Cracow  
Jean Ward, University of Gdańsk  
Jerzy Welna, University of Warsaw  
Florian Zappe, University of Göttingen

## GUEST REVIEWERS

Magdalena Bator, WSB University in Poznań  
Anna Cichosz, University of Łódź  
Kousuke Kaita, Chiba University, Japan  
Joanna Lewińska, University of Warsaw  
Andrzej Łęcki, Jan Kochanowski University in Kielce  
Monika Łodej, Jan Kochanowski University in Kielce  
Alisa Masiejczyk, University of Warsaw  
Agnieszka Piskorska, University of Warsaw  
Maciej Rataj, University of Gdańsk  
Maria Reraki, University of Birmingham  
Paweł Rydzewski, University of Warsaw



UNIVERSITY  
OF WARSAW

# TABLE OF CONTENTS

## **David L. White**

The Form of DO Employed to Form the Weak Preterit . . . . . 5

## **Aleksandra Kalaga**

The Semantics of Morphological Conversion in Old English . . . . . 33

## **Gabriela Brůhová and Kateřina Vašků**

Lexical Bundles Ending in *that* in Academic Writing by Czech Learners  
and Native Speakers of English . . . . . 53

## **Dorota Watkowska**

Redundancy in ELF: A Corpus-Based Study on Negative  
and Modal Concord . . . . . 71

## **Mayowa Akinlotan and Ayo Ayodele**

Discursive Chain and Movement in Crisis-Driven Nigerian Political  
Discourse: Corpus Evidence from Herdsmen Newspaper Headlines . . . . . 87

## **Anita Buczek-Zawiła**

Phonological Awareness of L1 Systemic Segmental Contrasts among  
Advanced ESL Speakers with Varied L1 Backgrounds . . . . . 107

## **Agnieszka Kaldonek-Crnjaković**

Development of Metalinguistic Awareness in EFL Vocabulary and Spelling:  
A Longitudinal Case Study of a Child and Adult with Dyslexia . . . . . 127

Gabriela Brůhová

 <http://orcid.org/0000-0001-5014-7965>

Charles University, Prague

Kateřina Vašků

 <http://orcid.org/0000-0002-0169-082X>

Charles University, Prague

## Lexical Bundles Ending in *that* in Academic Writing by Czech Learners and Native Speakers of English

---

**Abstract:** The aim of this paper is to explore how Czech learners of English use lexical bundles ending in *that* in their academic texts in comparison with novice and professional L1 authors. The analysis is based on three corpora (*VESPA-CZ*, *BAWE* and our own corpus of papers published in academic journals). The results suggest that Czech learners of English do not use a more limited repertoire of lexical bundles ending in *that* than professional writers. However, there are differences between the groups studied, especially in the range of various shell nouns used in nominal bundles. Novice writers, both L1 and L2, use bundles ending in *that* to express stance more frequently than professional writers.

**Keywords:** learner corpus, *VESPA*, lexical bundles, *that*, stance, academic writing

---

### 1. Introduction

In the last several decades, increasing attention has been paid to the study of various kinds of recurrent multi-word sequences. It is well-known that language “is composed of multi-word prefabricated expressions” (Biber et al. 2004, 372) and their use has been considered “a marker of proficient language use of a particular register, including academic writing” (Cortes 2004, 398). Multi-word sequences have been studied under various labels, e.g. lexical bundles (Biber et al. 1999; Cortes 2004; Chen and Baker 2010), clusters (Scott 1996), recurrent word combinations (Altenberg 1998), or n-grams (Granger and Bestgen 2014; Rayson 2015). In the present paper, the recurrent multi-word sequences are referred to as lexical bundles.

Lexical bundles can be defined as “bundles of words that show a statistical tendency to co-occur” or “sequences of words that most commonly co-occur in a register” (Biber et al. 1999, 989). Most bundles represent incomplete grammatical structures and are not idiomatic in meaning, e.g. *I don't know what, in the case of the, it should be noted that*, and they function as “basic building blocks of discourse” (Biber et al. 2004, 371). As pointed out by Biber et al. (1999, 992), “[t]o qualify as a lexical bundle, a word combination must frequently recur in a register”, the minimum frequency being ten times per million words in a register and their occurrences must be spread across at least five different texts.

The aim of this paper is to explore how Czech learners of English use lexical bundles ending in *that* in their academic texts compared to native speakers. The analysis was carried out on three corpora in order to find similarities and differences in the use of lexical bundles at different levels of writing proficiency, identifying patterns of learner overuse and/or underuse<sup>1</sup> of lexical bundles. The first corpus contains writing from Czech advanced learners of academic English using the Czech component of the *Varieties of English for Specific Purposes* database (*VESPA*). The two other corpora contain L1 writing: one from English L1 novice authors of academic texts using the *British Academic Written English* corpus (*BAWE*) and the other from English L1 professional writers of academic texts (using our own corpus of papers published in academic journals). The use of the three corpora enables us to study lexical bundles in two dimensions. Firstly, we compare L2 with L1 writers (both novice and professional) and secondly, we aim to investigate the differences between novice (both L1 and L2) and professional writers of academic texts. We assume that learners will make less use of multi-word expressions in their academic texts than native writers and that they will rely more heavily on the open-choice principle, i.e. they tend to combine individual words. L1 users, on the other hand, are expected to make more use of semi-preconstructed phrases, employing the idiom principle (for further discussion of open-choice and idiom principle, cf. Sinclair 1991).

## 2. Some previous studies in the field of L2 phraseology

Previous research has shown that the use of lexical bundles “unmistakably distinguishes native speakers of a language from L2 learners” (Granger and Bestgen 2014; cf. also Pawley and Syder 1983; Ebeling and Hasselgård 2015), and that the frequent and appropriate use of lexical bundles can be considered a sign of phraseological competence within a register, including academic writing. The fact that the use of appropriate multi-word expressions should be regarded as a marker of proficient language use of academic writing has been pointed out by Haswell (1991, 236), who claims that “as writers mature they rely more and more on collocations and that the lesser use of them accounts for some characteristic behaviour of apprentice writers”.

Some authors have focused on comparisons between native expert and native

student writing (Cortes 2004; Hyland 2008a). It has been demonstrated by Cortes (2004, 398) that “expert disciplinary writing differs from that of novices with respect to the use of discipline-specific frequent word combinations”. Furthermore, Cortes (2004, 398) has suggested that acquiring lexical bundles which are unique to a particular register is crucial not only for non-native, but also for native students. According to Granger (2017, 9), academic lexical items “represent a particularly significant hurdle for L2 users, who have to understand and produce academic language in a language that is not their own”.

As has been noted by several researchers, L2 writers tend to use a more restricted repertoire of lexical bundles than L1 writers, making use of the same bundles more frequently and in situations where L1 writers would opt for a different expression (cf. Chen and Baker 2010; Ädel and Erman 2012, 90; Garner 2016, 33; Vašků, Brůhová and Šebestová 2019).

In addition, L2 users seem to be less confident when writing in a foreign language, and as a result they “regularly clutch for the words [they] feel safe with” (Hasselgren 1994, 237). Hasselgren (1994) uses the teddy bear metaphor to describe the situation in which L2 learners tend to overuse familiar words. She proposes (1994, 250) that it is especially direct L1 transfer that gives rise to lexical teddy bears. Secondly, some teddy bears arise from perceived equivalence between L1 and L2. A third type of teddy bears is represented by expressions used in a context where native speakers would opt for another, synonymous expression. The metaphor of teddy bear was later transferred to multi-word expressions by Ellis (2012, 37), who explains that “phrasal teddy bears” are formulaic expression with routine functional purposes. Hasselgård (2019, 340) labels these multi-word units as “phraseological teddy bears” and defines them as expressions which “learners use more frequently and in more contexts than native speakers do”. She also explores their use in the context and points out for example that “the learners may have a tendency to over-express contrastive relations when the discourse moves from one topic to another” (2019, 351).

### 3. Material and method

As mentioned above, the study is based on material from three corpora. The first corpus is the Czech component of the international *VESPA* corpus of advanced learners’ English (henceforth *VESPA-CZ*). The Czech *VESPA* currently contains English texts written by Czech BA students during their English literature classes. The second corpus is the English Literature section of the *BAWE* corpus (henceforth *BAWE-EL*), which comprises L1 university students’ assignments. The third corpus is compiled from papers published in English literary academic journals, written by professional authors whose native language is English (henceforth *AP*). The latter two corpora are approximately twice the size of the *VESPA-CZ* corpus (see Table 1).

**Table 1.** Corpora used for the study

Corpus	Acad. Level	Source	Time	Register	Topic	size: tokens	no. of texts
VESPA-CZ	L2 novice	Charles University, Prague, Faculty of Arts, English Studies Programme, BA 2nd year students' essays	2016-2019	students' essays	English literature: Renaissance to Restoration	106 600	48
BAWE-EL	L1 novice	Universities of Warwick, Reading and Oxford Brookes, Arts and Humanities – English, good-standard students' essays	2004-2007	students' essays	English literature	236 000	89
AP	L1 professional	academic journals English Literary Renaissance, Renaissance Studies, Shakespeare Quarterly	1978-2014	academic papers	English literature: Renaissance	235 000	34

From these corpora, we retrieved all 3-word to 5-word bundles having *that* as their final element using AntConc (version 3.5.8, Anthony 2019). Since we searched for word forms of *that*, the query returned two functions of *that*, i.e. the subordinative conjunction and the relative pronoun.<sup>2</sup> In order to reflect the different sizes of the corpora, the raw cut-off frequency was set to two occurrences in *VESPA-CZ* and four occurrences in the L1 corpora distributed in at least two different texts. In the next step, all topic-specific bundles were excluded, namely those used to describe the storyline of the literary texts discussed, such as *to the reader that*, *the people that*, *the bond that*, as we focus on the general academic vocabulary and these are not likely to occur in other corpora.

In the analysis, we first identified the most frequent lexical bundles in each of the three corpora, focusing on the similarities and differences in their frequency and use (chi-square test was used to analyse the differences; the comparison has been made relative to the total number of words in the corpora). Next, lexical bundles in the sample were classified according to their structure and function. From the structural point of view, the bundles were categorized according to the word-class of the element preceding *that*, e.g. N *that*, ADJ *that*. The functional classification of the bundles is based on the classification proposed by Biber et al. (2004, 384), who distinguish the following three functional types:

- i) **Stance expressions** “express attitudes or assessments of certainty that frame some other proposition”, e.g. *are more likely to*, *it is necessary to*;
- ii) **Referential expressions** “make direct reference to physical or abstract entities, or to the textual context itself”, e.g. *is one of the*, *the beginning of the*, *of the things that*;
- iii) **Discourse organizers** “reflect relationships between prior and coming discourse”, e.g. *in this chapter we*, *on the other hand*, *if we look at*.

#### 4. Analysis and results

In the first step of the analysis we focused on the most frequent lexical bundles in the three corpora. The ten most frequent lexical bundles are presented in Table 2,<sup>3</sup> which also specifies their raw and relative frequency and their dispersion across texts. Two bundles proved to be prominent in all three corpora, namely *the fact that* and *the idea that* (marked in bold). What is, however, noteworthy is that the relative frequency of *the fact that* in the L2 corpus is considerably higher (37.5 per 100k tokens) than in the L1 corpora and there is a steeper decline in frequencies of the following bundles. This is in accordance with Hasselgård’s results, who notes that “learners tend to re-use a small number of bundles to a greater extent than native speakers” (Hasselgård 2019, 347). This tendency is evident especially in comparison with *BAWE-EL*. We can therefore assume that *the fact that* can be described as a “phraseological teddy bear” of Czech L2 users of academic





English, at least in terms of the frequency of use. Interestingly, one of the lexical bundles found in *BAWE-EL*, *be argued that*, could be regarded as a phraseological teddy bear of L1 novice writers as it is overused in *BAWE-EL*. Four bundles: *the fact that*, *be said that*, *by saying that* and *it is obvious that* are significantly (chi-square,  $p < 0.05$ ) overused by Czech learners of English when compared to both L1 corpora. *It seems that* is overused in both novice corpora (compared to the professional writers). On the other hand, several bundles are underused in novice texts (although below the level of statistical significance). They include e.g. *a way that*, *by suggesting that*, *it is that*.

Apart from the frequency of occurrence, the table also includes text dispersion, i.e. the range of texts in which the bundles occur. Although Hasselgård (2019, 359) suggests that “text dispersion may be a better indicator than frequencies per 100,000 words of over- and underuse of lexical bundles” in that “[the] most common bundles turned out to occur in a greater proportion of the texts in L1 English; learners are thus less uniform in their use of most of the frequent bundles”, our analysis of text dispersion does not show any considerable differences between the three corpora.

The bundles in the three samples were then classified based on their structure. Depending on the word-class of the element preceding *that*, we identified four main structures: the nominal type (N *that*), the verbal type (V *that*), the adjectival type (ADJ *that*) and the prepositional type (prep N *that*). Table 3 summarizes the structural categories identified in the three corpora.

**Table 3.** Overview of structural types

formal pattern	VESPA-CZ			BAWE-EL			AP		
	types freq.	tokens freq.	Tokens %	types freq.	tokens freq.	tokens %	types freq.	tokens freq.	tokens %
N <i>that</i>	9	67	32.8	14	157	34.6	7	87	43.7
V <i>that</i>	25	73	35.8	26	213	46.9	12	60	30.2
be ADJ <i>that</i>	10	28	13.7	6	48	10.6	2	8	4.0
prep N <i>that</i>	6	20	9.8	5	36	7.9	8	44	22.1
Other	7	16	7.8	0	0	0	0	0	0
<b>Total</b>	<b>57</b>	<b>204</b>	<b>100.0</b>	<b>52</b>	<b>454</b>	<b>100.0</b>	<b>30</b>	<b>199</b>	<b>100.0</b>

Contrary to our expectations, the analysis revealed that professional writers produce fewer lexical bundles ending in *that* and use them less frequently (chi-square test,  $p < 0.05$ ) than novice academic writers (both L1 and L2). While novice L2 and L1 writers produce 57 and 52 different lexical bundles (i.e. types), respectively, professional L1 writers employ only 30 lexical bundles. Similarly, the

overall frequency of bundles in the professional corpus (i.e. tokens) represents only a half of the tokens in the novice corpora.<sup>4</sup>

As regards the structural classification, the nominal and verbal types are the most frequent in all three corpora. When the overall frequencies of bundles are taken into account, the verbal type is slightly more frequent than the nominal type in novice academic texts (73 vs. 67 occurrences in *VESPA-CZ* and 213 vs. 157 occurrences in *BAWE-EL*), while in the *AP* corpus, the nominal type prevails (87 nominal bundles vs. 60 verbal ones). The adjectival type seems to be used more by novice writers, especially L2 writers (10 types/28 tokens in *VESPA-CZ* vs. 2 types/8 tokens in *AP*). On the other hand, professional writers use the widest range of prepositional sequences (8 different sequences). In addition to the four basic structural types, the *VESPA-CZ* also contains a small number of other bundles, including *more than that*, *but since that*, *not only that*.

The following subsections describe the four structural types individually, comparing the results in detail.

#### 4.1 N *that* bundles

As mentioned above, N *that* bundles were frequent in all three corpora. All instances of these bundles are presented in Table 4. The greatest diversity of nominal bundles is to be found in L1 novice academic texts and the lowest diversity in texts written by professional authors. Note that N *that* bundles include two syntactic constructions, namely nouns followed by a nominal content clause introduced by the conjunction *that* (1a) and nouns followed by a relative clause introduced by the relative pronoun *that* (1b):

- (1a) Again, the speaker represents the idea **that one's spirit does not have to die** when the body does but can exist longer.  
 (1b) Time brings change and if there is one thing **that is unlikely to change** it is precisely that.

The investigation of the meaning of the nouns within this pattern reveals that there is a significant group of nouns belonging to the same category, which can be described as shell nouns.<sup>5</sup> A shell noun is defined as “potentially any abstract noun, the meaning of which can only be made specific by reference to its context” (Flowerdew 2003, 329). The category includes nouns such as *fact*, *idea*, *belief*, *notion*, which function as “empty shells” in certain contexts “because they can enclose or anticipate the surrounding discourse” (Aktas and Cortes 2008, 4). Ivanič (1991, 96) suggests that “since [shell nouns] take on the bulk of their meaning from context, they are not subject-specific. This makes them particularly useful for learners entering the academic discourse community”.

As can be seen in examples (1a), (2a) and (2b), the shell nouns used in N *that*

**Table 4.** N *that* bundles

VESPA-CZ	raw freq	per 100k	BAWE-EL	raw freq	per 100k	AP	raw freq	per 100k
<b>the fact that</b>	407	37.5	<b>the fact that</b>	68	28.8	<b>the fact that</b>	48	20.4
<b>the idea that</b>	4	6.6	<b>the idea that</b>	24	10.2	<b>the idea that</b>	11	4.7
is the fact that	4	3.8	<b>the notion that</b>	8	3.4	a way that	9	3.8
one thing that	3	3.8	the belief that	7	3.0	the way that	6	2.6
but the fact that	3	2.8	the sense that	7	3.0	<b>the notion that</b>	5	2.1
<b>the notion that</b>	2	2.8	the effect that	6	2.5	is the fact that	4	1.7
the conclusion that	2	1.9	the impression that	6	2.5	the assumption that	4	1.7
the only thing that	2	1.9	a concept that	5	2.1			
the thought that		1.9	a way that	5	2.1			
			the suggestion that	5	2.1			
			the conclusion that	4	1.7			
			the feeling that	4	1.7			
			the view that	4	1.7			
			the way that	4	1.7			
<b>Total</b>	<b>67</b>	<b>63</b>		<b>157</b>	<b>66.5</b>		<b>87</b>	<b>37</b>

bundles anticipate the following *that*-clause, which carries the meaning and provides more information about the entity. Since the most common shell nouns (*fact*, *idea*) are semantically empty and can be regarded as somewhat redundant, their overuse contributes to “the impression of verbosity” (Granger 1998, 156) in L2 writing.

- (2a) The ultimate irony lies in the **fact** that he knows what the lady had done. (*VESPA-CZ*)
- (2b) the ‘essence’ of structuralism is the **belief** that things cannot be understood in isolation (*BAWE-EL*)

As regards the function of N *that* bundles, they typically “make direct reference to (...) abstract entities” (Biber et al. 2004, 384), and therefore their function can be described as referential. However, some of the nouns border on the category of stance bundles, expressing certainty (*fact*) or a lower degree of certainty (*belief*, *impression*, *assumption*) or on the category discourse organizers (*conclusion*). In addition, *the fact that* can be a part of a longer bundle *due to the fact that* (equivalent to *because*), which is a discourse organiser.

#### 4.2 V *that* bundles

The category of V *that* bundles exhibits the highest diversity, especially in novice writing. This confirms Chen and Baker’s results that novice writers use more verb-based bundles than native professional writers (cf. Chen and Baker 2010, 36). There are only two bundles common to all three corpora: *to say that* and *to note that*.

Table 5. *V that bundles*

VESPA-CZ	raw freq	per 100k	BAWE-EL	raw freq	per 100k	AP	raw freq	per 100k
be said that	6	5.6	be argued that	30	12.7	to suggest that	7	3.0
it seems that	5	4.7	to suggest that	22	9.3	it is that	6	2.6
to say that	5	4.7	it could be argued that	17	7.2	points out that	6	2.6
he claims that	4	3.8	to say that	12	5.1	to say that	6	2.6
would suggest that	4	3.8	he argues that	11	4.7	he admits that	5	2.1
be argued that	3	2.8	to argue that	11	4.7	I suggest that	5	2.1
can be said that	3	2.8	it seems that	10	4.2	point is that	5	2.1
he argues that	3	2.8	points out that	9	3.8	he suggests that	4	1.7
he states that	3	2.8	to believe that	9	3.8	is not that	4	1.7
interesting to note that	3	2.8	it might be argued that	7	3.0	reminds us that	4	1.7
might be that	3	2.8	might suggest that	7	3.0	to argue that	4	1.7
to believe that	3	2.8	he believes that	6	2.5	to note that	4	1.7
to note that	3	2.8	to show that	6	2.5			
to point out that	3	2.8	is not to say that	5	2.1			
be shown that	2	1.9	this suggests that	5	2.1			
he knows that	2	1.9	to explain that	5	2.1			
he realizes that	2	1.9	to note that	5	2.1			
is interesting to note that	2	1.9	be suggested that	4	1.7			
it can be said that	2	1.9	can see that	4	1.7			
it may seem that	2	1.9	he felt that	4	1.7			
not remember that	2	1.9	is interesting to note that	4	1.7			
point here is that	2	1.9	is possible to suggest that	4	1.7			
pointing out that	2	1.9	it appears that	4	1.7			
this means that	2	1.9	possible to argue that	4	1.7			
to realize that	2	1.9	reminds us that	4	1.7			
			to understand that	4	1.7			
<b>Total</b>	<b>73</b>	<b>68.7</b>		<b>213</b>	<b>90.2</b>		<b>60</b>	<b>25.6</b>

The heterogeneity of the lists is given by the fact that the English verb phrase may be very complex, expressing grammatical categories such as tense or voice and it may be part of various syntactic structures.

The analysis revealed that V *that* bundles can be subclassified into the following four syntactic structures:

- i) **personal subject + active verb + *that***: *he argues that, he believes that, he suggests that*;
- ii) ***to*-infinitive + *that***: *to note that, to realize that, to show that, to argue that*;
- iii) **(anticipatory *it*) + (modal verb) + passive verb + *that***: *it can be said that, it might be argued that, it could be argued that*;
- iv) **(anticipatory *it*) + (copular verb) + adjective + *to*-infinitive + *that***: *is interesting to note that, is possible to suggest that, possible to argue that*.

Many of the verbs used in V *that* bundles can be characterized as reporting verbs (e.g. *argue, say, suggest, note*). These verbs are typical of the academic discourse, since their function is to refer to another author's work or to introduce someone's opinions, ideas or assumptions (cf. Hyland 1998). The typical structure used for the reporting function is the first subtype in which the verb is in the active voice and follows a personal subject (*he argues that*). In these cases, the function of lexical bundles proved to be difficult to determine due to the fact that they represent borderline cases between referential expressions and discourse organizers. On the one hand, they make direct reference to entities and, on the other hand, they help to organize text because they are used to introduce a specific feature of academic texts, i.e. reference to other authors.

Apart from the above-mentioned functions, a large number of V *that* bundles function as stance expressions. "Stance bundles provide a frame for the interpretation of the following proposition, conveying two major kinds of meaning: epistemic and attitude/modality" (Biber et al. 2004, 389-390). Epistemic stance bundles express certainty or uncertainty, while attitudinal stance bundles express the speaker's attitude towards the actions described in the following proposition. Most bundles in our corpora express epistemic modality (e.g. *it seems/appears that, would suggest that, might be that, it may seem that, it could be argued that, is possible to suggest that*); attitudinal modality is rarer (*is interesting to note that*). Again, the writers use these bundles as hedges "introducing a degree of tentativeness to what is being reported" (Biber et al. 2004, 410). Very often, typically with the anticipatory *it* (type iii. and iv.), stance or evaluation are expressed impersonally (cf. impersonal stance bundles in Biber et al. 2004, 389; Hyland 2008b, 18). Focusing on the use of the passive in the three corpora, we found that the list retrieved from the professional texts does not contain any recurrent passive structures (type iii.); novice authors, on the other hand, tend to use these bundles more frequently. It seems that although the passive is generally regarded as being typical of the academic discourse, the results from the L1 professional corpus imply that the passive does not occur frequently in any particular bundle. The passive

used with some specific verbs is clearly overused by both L1 and L2 novice writers (*it can/could/might be argued that, it can be said that*) and these bundles can be considered as phraseological teddy bears of novice academic writers. The use of modal verbs expressing lack of certainty (*can, could, may, might*) also contributes to the function of hedging, communicating uncertainty or caution.

### 4.3 ADJ *that* bundles

The most common structure of ADJ *that* bundles is **anticipatory *it* + copula + ADJ + *that***, where *that* is a part of an extraposed subject clause. As can be seen from Table 6, the only bundle shared by all three corpora is ***it is clear that*** (marked in bold).

**Table 6.** ADJ *that* bundles

VESPA-CZ	raw freq	per 100k	BAWE-EL	raw freq	per 100k	AP	raw freq	per 100k
<b>it is clear that</b>	5	4.7	<b>it is clear that</b>	10	4.2	<b>it is clear that</b>	4	1.7
it is obvious that	5	4.7	becomes clear that	10	4.2	it is significant that	4	1.7
becomes apparent that	3	2.8	it is possible that	9	3.8			
it is true that	3	2.8	it becomes clear that	8	3.4			
is aware that	2	1.9	it is interesting that	7	3.0			
it becomes apparent that	2	1.9	it is likely that	4	1.7			
it is evident that	2	1.9						
it is only natural that	2	1.9						
makes it clear that	2	1.9						
similar to that	2	1.9						
<b>Total</b>	<b>28</b>	<b>26.4</b>		<b>48</b>	<b>20.3</b>		<b>8</b>	<b>3.4</b>

It is striking that the professional corpus contains only two ADJ *that* bundles, whereas novice writers obviously show preference for this type. From the functional point of view, these bundles can be classified as stance bundles (similarly to V *that* bundles). Our findings indicate that the majority of ADJ *that* bundles are used to express epistemic modality, conveying either the writer's certainty (e.g. *it is clear that, it is obvious that, it is evident that*), or uncertainty (*it is possible that, it is likely that*). In addition, ADJ *that* bundles can also express the speaker's attitude (*it is only natural that, it is interesting that, and it is significant that*). As regards the epistemic modality, the bundles expressing uncertainty are – contrary to our expectations – not as frequent as those expressing certainty, and they only appear in the L1 novice corpus. This corroborates the findings of Chen and Baker (2010, 43), who suggest that L2 writers manifest “[t]he tendency to hedge less and instead adopt an overstating tone”. This tendency “seems to be universal for learners from different L1 backgrounds” (2010, 43). A similar idea has been pointed out by Hewings and

Hewings (2002, 380-381), who claim that “student writers make a much greater and more overt effort to persuade readers of the truth of their statements than do the published writers”. It should be noted that even stance bundles expressing a high degree of certainty can be viewed as a means of hedging as they allow writers “to present information as an opinion rather than accredited fact” (Hyland 2005, 178).

#### 4.4 Prep N *that* bundles

Two prep N *that* bundles were found in all three corpora: *to the fact that* and *by the fact that*. Especially in novice writing, most prep N *that* bundles are extensions of the nominal type containing a shell noun: *fact* or *idea*. Our results suggest that L2 Czech learners overuse the bundle *by saying that*, whereas native professional writers opt for other bundles, using more sophisticated verbs to express the same meaning, namely *by suggesting that* and *by claiming that*. However, due to the low number of tokens and their nature, no conclusive results can be drawn from the analysis of prep N *that* bundles.

**Table 7.** Prep N *that* bundles

VESPA-CZ	raw freq	per 100k	BAWE-EL	raw freq	per 100k	AP	raw freq	per 100k
by saying that	5	4.7	<b>to the fact that</b>	14	5.9	<b>by the fact that</b>	10	4.3
<b>to the fact that</b>	5	4.7	due to the fact that	9	3.8	in a way that	7	3.0
<b>by the fact that</b>	4	3.8	with the idea that	5	2.1	by suggesting that	6	2.6
as something that	2	1.9	<b>by the fact that</b>	4	1.7	of all that	5	2.1
due to the fact that	2	1.9	in the sense that	4	1.7	by claiming that	4	1.7
in the fact that	2	1.9				in ways that	4	1.7
						to the extent that	4	1.7
						<b>to the fact that</b>	4	1.7
<b>Total</b>	<b>20</b>	<b>18.9</b>		<b>36</b>	<b>15.2</b>		<b>44</b>	<b>18.8</b>

## 5. Conclusions

The paper has explored how Czech learners of English use lexical bundles ending in *that* in academic texts compared to native speakers. The analysis has revealed both similarities and differences in the two dimensions studied, i.e. between L1 and L2 writing and between novice and professional writing.

In contrast with our expectations, Czech learners of English do not use a more limited repertoire of lexical bundles ending in *that*. Surprisingly, it is the professional L1 writers whose repertoire is most restricted. This suggests that the study of a specific structural pattern (bundles ending in *that* in our case) may yield different results than the study of lexical bundles in academic texts in general. Thus, the

structures containing a *that*-clause seem to be favoured by authors who are in the process of learning academic language and their excessive use by novice writers may create the impression of verbosity. In addition, novice writers, struggling to master the intricacies of the academic register, rely more strongly on selected multi-word sequences which they are comfortable with (i.e. phraseological teddy bears, e.g. *the fact that, be said that, by saying that, it is obvious that*).

As regards the formal classification, all four types (N *that*, V *that*, ADJ *that* and prep N *that*) are represented in the three corpora, with the nominal and verbal type being most frequent. Novice writers tend to overuse the adjectival type, while professional writers use the widest range of prepositional sequences. An important semantic subtype of N *that* bundles is represented by bundles containing a shell noun. The results suggest that novice writers, both L1 and L2, are aware of their use in the academic register. However, there are differences between the two groups of novice writers in that L1 novice writers use a wide range of shell nouns with various meanings (*notion, belief, impression, suggestion, etc.*), whereas L2 novice writers show preference for a limited number of generally applicable nouns (*fact, idea, thing*). L1 professional writers, again, use a restricted range of N *that* bundles (i.e. types).

The functional analysis revealed that the function of lexical bundles ending in *that* is clearly associated with their structural type. While N *that* bundles were found to perform primarily the referential function, V *that* bundles, being mostly used with reporting verbs, can be viewed as borderline cases between referential and discourse organising bundles, and ADJ *that* bundles generally express stance. The area in which the texts differed most was the expression of stance. Our findings indicate that novice writers use bundles ending in *that* to express stance more frequently than professional writers. Most stance bundles express epistemic modality. The bundles were used not only to express uncertainty or caution, i.e. as a means of hedging, but also to express certainty. We argue that the expression of certainty can be seen as a means of overstatement, possibly used with the intention to persuade the reader of the writer's opinion.

The results imply that stance is typically expressed by structures employing the anticipatory *it*, both in V *that* bundles with a passive verb and ADJ *that* bundles. The anticipatory *it* and the passive voice allow the writer to express stance impersonally without the attitude being directly attributed to the author (cf. Hyland 2008b, 18). It is curious that novice authors were found to use the passive structures more frequently than professional writers. Although the passive is generally considered to be a characteristic feature of the academic register, this particular structure, i.e. (anticipatory *it*) + (modal verb) + passive verb + *that*, does not occur among the lexical bundles retrieved from L1 professional texts.

We hope to have shown that in some aspects, there are no significant differences between the language of L2 novice academic writers and L1 novice academic writers, who both struggle with acquiring academic proficiency, while in other aspects there is a clear distinction between the language used by L2 and L1 writers



regardless of their proficiency. If we were to propose a direction for future research, we would choose to focus on the use of reporting verbs in order to investigate the range and sophistication of these verbs in L2 academic texts.

## Notes

- 1 As has been pointed out by Paquot and Granger (2012, 143), “[in learner corpus research], the terms overuse and underuse are descriptive, not prescriptive terms: they refer to the fact that a linguistic feature is found significantly more or less often in a learner corpus than in a reference native or expert corpus”.
- 2 No instance of the demonstrative *that* has been found in the bundles retrieved.
- 3 The tenth position is occupied by all the bundles with the threshold frequency.
- 4 Note that *VESPA-CZ* is approximately half the size of the other two corpora.
- 5 Shell nouns have been referred to by various terms: carrier nouns (Ivanič 1991), general nouns (Halliday and Hasan 1976), signalling nouns (Flowerdew 2003).

## References

- Ädel, Annelie, and Britt Erman. 2012. “Recurrent Word Combinations in Academic Writing by Native and Non-native Speakers of English: A Lexical Bundles Approach.” *English for Specific Purposes* 31.2: 81–92.
- Aktas, Rahime Nur, and Viviana Cortes. 2008. “Shell Nouns as Cohesive Devices in Published and ESL Student Writing.” *Journal of English for Academic Purposes* 7.1: 3–14.
- Altenberg, Bengt. 1998. “On the Phraseology of Spoken English: The Evidence of Recurrent Word-combinations.” Ed. A.P. Cowie. *Phraseology, Theory, Analysis, and Applications*. Oxford: Oxford University Press. 101–124.
- Anthony, Laurence. 2019. *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/>.
- Biber, Douglas., Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. “If You Look at...: Lexical Bundles in University Lectures and Textbooks.” *Applied Linguistics* 25: 371–405.
- Chen, Yu-Hua, and Paul Baker. 2010. “Lexical Bundles in L1 and L2 Academic Writing.” *Language Learning & Technology* 14.2: 30–49.
- Cortes, Viviana. 2004. “Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology.” *English for Specific Purposes* 23: 397–423.
- Ebeling, Signe Oksefjell, and Hilde Hasselgård. 2015. “Learner Corpora and

- Phraseology.” *The Cambridge Handbook of Learner Corpus Research*. Ed. Sylviane Granger, Gaëtanelle Gilquin, and Fanny Meunier. Cambridge: Cambridge University Press. 207–230.
- Ellis, Nick C. 2012. “Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear.” *Annual Review of Applied Linguistics* 32: 17–44.
- Flowerdew, John. 2003. “Signalling Nouns in Discourse.” *English for Specific Purposes* 22.4: 329–346.
- Garner, J.R. 2016. “A Phrase-framed Approach to Investigating Phraseology in Learner Writing across Proficiency Levels.” *International Journal of Learner Corpus Research* 2.1: 31–68.
- Granger, Sylviane. 1998. “Prefabricated Patterns in Advanced EFL Writing.” *Phraseology: Theory, Analysis and Applications*. Ed. A.P. Cowie. Oxford: Clarendon Press. 145–160.
- Granger, Sylviane. 2017. “Academic Phraseology. A Key Ingredient in Successful L2 Academic Literacy.” *Oslo Studies in Language* 9.3: 9–27.
- Granger, Sylviane, and Yves Bestgen. 2014. “The Use of Collocations by Intermediate vs. Advanced Non-native Writers: A Bigram-based Study.” *International Review of Applied Linguistics in Language Teaching* 52.3: 229–252.
- Halliday, Michael Alexander Kirkwood, and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Hasselgård, Hilde. 2019. “Phraseological Teddy Bears: Frequent Lexical Bundles in Academic Writing by Norwegian Learners and Native Speakers of English”. Ed. Viola Wiegand, and Michaela Mahlberg. *Corpus Linguistics, Context and Culture*. Berlin: De Gruyter. 339–362.
- Hasselgren, Angela. 1994. “Lexical Teddy Bears and Advanced Learners: A Study into the Ways Norwegian Students Cope with English Vocabulary.” *International Journal of Applied Linguistics* 4.2: 237–258.
- Haswell, Richard H. 1991. *Gaining Ground in College Writing: Tales of Development and Interpretation*. Texas: Southern Methodist University Press.
- Hewings, Martin, and Ann Hewings. 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.
- Hyland, Ken. 1998. *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.
- Hyland, Ken. 2005. “Stance and Engagement: A Model of Interaction in Academic Discourse.” *Discourse Studies* 7.2: 173–191.
- Hyland, Ken. 2008a. “Academic Clusters: Text Patterning in Published and Postgraduate Writing.” *International Journal of Applied Linguistics*. 18.1: 41–62.
- Hyland, Ken. 2008b. “As Can Be Seen: Lexical Bundles and Disciplinary Variation.” *English for Specific Purposes* 27.1: 4–21.
- Ivanič, Roz. 1991. “Nouns in Search of a Context: A Study of Nouns with Both Open- and Closed-system Characteristics.” *IRAL-International Review of Applied Linguistics in Language Teaching* 29.2: 93–114.

- Paquot, Magali, and Sylviane Granger. 2012. "Formulaic Language in Learner Corpora." *Annual Review of Applied Linguistics*. 32.32: 130–149.
- Pawley, Andrew, and Frances Hodgetts Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." *Language and Communication*. 191–226.
- Rayson, Paul Edward. 2015. "Computational Tools and Methods for Corpus Compilation and Analysis." *Cambridge Handbook of English Corpus Linguistics*. Ed. Douglas Biber, and Randi Reppen. Cambridge: Cambridge University Press. 32–49.
- Scott, Mike. 1996. *Wordsmith Tools 4*. Oxford: Oxford University Press.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Vašků, Kateřina, Gabriela Brůhová, and Denisa Šebestová. 2019. "Phraseological Sequences Ending in *of* in L2 Novice Academic Writing." *International Conference on Computational and Corpus-Based Phraseology*. Springer: Cham. 431–443.

## Corpora

*BAWE (British Academic Written English Corpus)*

<https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/>

*VESPA-CZ (Varieties of English for Specific Purposes Database)*

<https://uclouvain.be/en/research-institutes/ilc/cecl/vespa.html>

## Acknowledgements

This study is a result of the research funded by the Czech Science Foundation as the project GA CR 19-05180S "Phraseology in English academic texts written by Czech advanced learners: a comparative study of learner and native speaker discourse".

GABRIELA BRŮHOVÁ is Senior Lecturer in the Department of English Language and ELT methodology at Charles University in Prague. Her research interests include phraseology, learner Corpora, syntax and information structure.

KATEŘINA VAŠKŮ is Senior Lecturer in the Department of English Language and ELT methodology at Charles University in Prague. Her research interests relate primarily to phraseology, learner corpora, word-formation and lexical semantics.