

Donatella Malavasi

Metadiscursive Constructions in Linguistics Essays by Italian EFL vs. L1 English University Students

Abstract

In the field of academic discourse analysis close attention has been paid to the study of metadiscourse and to the analysis of recurrent sequences of words, variously called phraseology, lexical bundles or formulaic language. In particular, multi-word units have been comparatively explored in writing by both native vs. non-native speakers of English, and novice vs. expert authors. Although the native/non-native divide has been questioned in favour of novice vs. expert distinction and expertise as more important aspects affecting language patterning, extensive research has considered academic writing by native English speakers vs. English as a Foreign Language (EFL) learners. Based on the assumption that native student writers and English learners have different levels of literacy in English and do not face the same difficulties in utilising formulaic language, the present study sets out to investigate the metadiscursive constructions most frequently used in academic writing by Italian EFL vs. L1 English students. The non-native material consists of theoretical and empirical essays in English Linguistics and Translation written by third-year students graduating in Foreign Languages at an Italian University. With regard to the native-speaker material for comparison, part of the British Academic Written English (BAWE) corpus was utilised, and essays written by L1 English students of linguistics were selected. With the support of corpus linguistic tools, recurrent metadiscursive bundles were identified in the two databases and examined both quantitatively and qualitatively. The analysis points to and discusses aspects of convergence and divergence between Italian EFL and L1 English student essays in the use of metadiscursive constructions.

Keywords: *metadiscourse, metadiscursive bundles, academic writing, essays, corpus analysis*

1. Introduction

In the field of academic discourse analysis, extensive research has been carried out on different genres (Carrió-Pastor 2020; Kuteeva and Mauranen 2018; Flowerdew 2014; Berkenkotter, Bhatia and Gotti 2012; Hyland and Sancho Guinda 2012; Hyland 2009; Hyland and Bondi 2006; Bhatia 2004, 1993; Swales 2004, 1990) as well as their metadiscursive features (see, for example, Bamford and Bondi 2005; Hyland 2005, 1998; Dahl 2004; Mauranen 1993).

Specifically in the area of metadiscourse, which is traditionally intended as discourse about discourse (Crismore, Markkanen and Steffensen 1993; Vande Kopple 1985), different strands of research have contributed to the understanding of this concept. While from narrower perspectives, metadiscourse is understood as metatext and text reflexivity (Mauranen 1993), or as metatext and writer-reader interaction (Ädel 2006), broad approaches to metadiscourse consider to be metadiscursive all linguistic resources that signal the presence of a text-organising, content-evaluating, and reader-engaging author (Hyland 2017, 2005; Hyland and Tse 2004). In particular, borrowing from Hyland's model, metadiscourse either fulfils the interpersonal function by making the writer visible through the expression of attitudes towards the message or interaction with readers (i.e., the interactional dimension of metadiscourse), or it serves the textual function by having the addresser enter the text to guide the addressee through the unfolding discourse (i.e., the interactive dimension of metadiscourse). More to the point, interactive metadiscourse concerns the ways writers "shape and constrain a text to meet the needs of particular readers, setting out arguments so that they will recover the writer's preferred interpretations and goals" (Hyland 2005, 49; 50-52). On the other hand, interactional material is used by writers to make their views explicit and to engage the readers as active participants in the text (see Hyland 2005, 52-54).

The increased attention to metadiscourse has been accompanied by a burgeoning interest in the analysis of recurrent sequences of words, variously called phraseology, lexical bundles, clusters and formulaic language (Durrant 2017; Simpson-Vlach and Ellis 2010; Granger and Meunier 2008; Biber and Barbieri 2007; Nesi and Basturkmen 2006; Biber, Conrad and Cortes 2004; Cortes 2004; Wray 2000; Biber et al. 1999). In the context of discourse analysis, multi-word sequences have been studied under many rubrics and from different perspectives, which all challenge the view that language is strictly compositional, arguing instead that it is highly formulaic and idiomatic (Sinclair 2004, 1991).

A particularly promising area has been research on lexical bundles or multi-word sequences investigated from an empirical, automated frequency-driven approach. From this perspective, lexical bundles are intended as continuous word sequences, which are semantically transparent and non-idiomatic in meaning, and which are "retrieved by taking a corpus-driven approach with specified frequency and distribution criteria" (Chen and Baker 2010, 30). According to the existing literature, recurrent sequences have been identified through corpus analysis that includes frequency thresholds and dispersion requirements. Firstly, the frequency threshold or cut-off frequency "determines the number of lexical bundles to be included in the analysis" (Chen and Baker 2010, 32). Albeit subjective and varying, frequency cut-offs of 25 times per

million words have been widely used in studies of written corpus data (e.g., Ädel and Erman 2012; Chen and Baker 2010) to identify bundles that recur enough to be considered as typical of a register or a set of texts. Secondly, the dispersion criterion is “the requirement that combinations have to occur in different texts, usually in at least 3-5 texts [...], which helps to avoid idiosyncrasies from individual writers/speakers” (Chen and Baker 2010, 32). Finally, as for the length of word combinations, four-word sequences are found to be the most researched length for writing studies (Ädel and Erman 2012; Chen and Baker 2010; Biber and Barbieri 2007; Cortes 2004). As suggested by Hyland (2008b, 8), four-word combinations “are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles.” Irrespective of their frequency and dispersion cut-offs, which are somewhat arbitrary,¹ lexical bundles serve important discourse functions in both spoken and written texts. On the basis of the functional taxonomy proposed by Hyland (2008a; 2008b), bundles can be grouped into three categories: research-, text-, and participant-oriented bundles. Specifically, research-oriented sequences of words “help writers to structure their activities and experiences of the real world,” text-oriented bundles are “concerned with the organisation of the text and its meaning as a message or argument,” and participant-oriented ones “are focused on the writer or reader of the text” (Hyland 2008b, 13-14).

Especially in academic writing, formulaicity, i.e., “knowledge of conventionalised multi-word combinations” (Pérez-Llantada 2014, 84), and lexical bundles have prominently been investigated from a learner corpus perspective (Römer, Cortes and Friginal 2020; Paquot and Granger 2012) to gain a better understanding of learners’ use of formulaic sequences. A wealth of studies has focused on the comparison of native vs. non-native students’ use of prefabricated expressions (Nam 2020; Güngör and Uysal 2016; Ädel and Erman 2012; Chen and Baker 2010), while others have explored phraseological constructions produced by novice vs. expert writers (Mbodj and Crossley 2020; Chen and Baker 2010; Granger and Paquot 2009; Shaw 2009; Hyland 2008a; Cortes 2004). Although the notion of ‘Native-speakerhood’ and the native/non-native dichotomy have been questioned as the main factor for explaining differences in academic writing and language patterning (Habibie and Hyland 2019; Hyland 2019; Römer 2009), the

¹ Frequency thresholds, which differ from one study to another, mainly depending on corpus size, have ranged from 10 (Biber et al. 1999) to 20 (Hyland 2008a, 2008b; Cortes 2004) and 40 times per million words (Biber, Conrad and Cortes 2004). As for the distribution criterion, researchers have taken into account sequences that occur at least in 3 to 5 texts (Biber and Barbieri 2007) or 10 percent of the texts (Hyland 2008a, 2008b). Finally, despite some ‘exceptions’ (see for example Esfandiari and Barbary 2017; Simpson-Vlach and Ellis 2010), four-word strings are the most extensively researched length (Hyland 2012, 151).

comparison of native and non-native writing remains a key research focus as it offers valuable insights into lexical bundle use.

From a language development approach, studies of lexical bundles, which are widely accepted to be “[a]n important component of fluent linguistic production” (Hyland 2012, 150), were inspired by the recognition that “to be a successful academic writer, an L2 learner is required to be competent at using these conventional sequences which characterize the learner’s discipline” (Li and Schmitt 2009, 86; Hyland 2008b). As a result, general consensus was reached among researchers in this field that L1 English writers use more lexical bundles and more varied bundle types than non-native English writers, who on the other hand tend to use less diverse formulaic sequences and overuse high-frequency clusters or the ones they master best. In view of these considerations, the present study sets out to investigate the use of English-language metadiscursive bundles in advanced learner writing by L1 speakers of Italian and in comparable L1 English writing, all produced by university students in the discipline of linguistics. Particular attention is paid to extended stretches of language which contribute to the organisation of the text, namely transition, resultative, structuring and framing signals, and to the expression of the writer’s stand on the issue(s) being discussed while pulling readers into the discourse, i.e., stance and engagement features (see Hyland 2008a, 49; 2008b, 14-19). The results of the comparative study of the metadiscursive constructions used by Italian learners of English as a Foreign Language (EFL) and L1 English students are discussed in Section 3 after a description of the materials under investigation and the methodological approaches adopted (Section 2). Finally, some considerations are drawn in the Discussion and conclusions section that throws light on aspects of convergence and divergence between learner and native speaker production in the use of metadiscursive chunks of language.

2. Corpora and methods

In an effort to contribute to the limited research available on the use of lexical bundles and in particular metadiscursive constructions by L1 Italian learners of English versus L1 English students, the present analysis was carried out based on two small yet comparable corpora. The first corpus collects 54 essays (approximately 171,400 tokens or running words) written in English by two groups of third-year Italian students graduating in Foreign Languages at an Italian University. This Bachelor’s program requires students to complete 180 credits (60 per year for 3 years) in different disciplinary areas, including foreign languages, linguistics, literature, economics and law. During their degree, students are required to study at least two foreign languages and reach a C1 level in one language, which is English for most students, and

a B2 level in the other. The students in the sample are third-year undergraduates and their competence in English can be considered upper-intermediate/advanced level. The writing represented in the so-called learner corpus is the students' final written productions for two courses: one in English linguistics and one in Italian/English translation. The essays collected are both theoretical and empirical, and were analysed in their original and not revised versions. The second database, which serves as the reference corpus, was drawn from an existing one, namely the British Academic Written English (BAWE) corpus, which contains approximately 3,000 pieces of proficient assessed student writing from British universities. To ensure comparability, only part of the BAWE corpus was used to represent native writing. In particular, 73 texts (ca. 171,800 tokens) written by L1 English students in the discipline of linguistics were selected for investigation.

The study involved a quantitative and qualitative analysis of extended units of language which contribute to the metadiscursive construction and characterisation of the essays under consideration. The research adopted a frequency-driven approach to multi-word units, based on the analysis of the most frequent word sequences. In particular, with the support of WordSmith Tools 6.0 (Scott 2012), four-word clusters were automatically extracted. Among the 15 most frequent multi-word units retrieved by the software, possible candidates for interactive and interactional metadiscourse were selected and irrelevant ones excluded. The frequency threshold for determining four-word lexical bundles was set to 25 times per million words (Ädel and Erman 2012; Chen and Baker 2010), and the dispersion criterion of five texts was used (Biber, Conrad and Cortes 2004). Afterwards, the metadiscursive sequences identified were explored in their wider context of use and categorised according to Hyland's (2008a; 2008b) functional taxonomies of text- and participant-oriented bundles. Thus, the present study is mainly informed by a broad approach to metadiscourse, which considers to be metadiscursive those four-word chunks that both refer to the current discourse, its addresser and addressee, and express the writer's attitude towards what is written. Corpus linguistic methods were combined with discourse analysis to investigate similarities and differences in the metadiscursive sequences produced in advanced learner academic writing by Italian EFL and L1 English novice authors.

3. Findings

For the purpose of this study, the metadiscursive bundle use in Italian EFL and L1 English essays was explored firstly quantitatively (see Subsection 3.1) and then qualitatively (see Subsection 3.2). Specifically, four-word metadiscursive sequences, selected as the scope of this

study, were quantitatively scrutinised in the essays written by native and non-native students in the discipline of linguistics. This analysis was then followed by an in-depth investigation of their functional characteristics.

3.1 Quantitative analysis of metadiscursive bundles

In order to understand how metadiscursive strings are used across the two corpora, the analysis examined the four-word clusters which were extracted by using WordSmith's function WordList cluster. Only the text- and participant-oriented constructions that fulfilled interactive and interactional metadiscourse functions and met the frequency and dispersion criteria described in Section 2 were taken into account. The sample of clusters retrieved from each set of student academic writing is listed in Table 1 together with their raw frequency of occurrence and the number of texts in which they occur.

A preliminary examination of the constructions featured in the table below shows that, although Italian EFL and L1 English university students produce a similar total number of bundles, the native-speaker writing contains a wider range of metadiscursive chunks than the non-native writing, with a total of 21, as compared to 14. These differences in bundle types suggest that native writers make use of more varied metadiscursive units, while Italian students tend to resort to a more restricted repertoire of constructions. This finding confirms observations made in previous studies on the tendency of EFL learners to opt for “phrasal teddy bears” in Ellis' words (2012) or repeated conventional patterns that are safe and reliable forms of expression. However, it is to be pointed out that the fact that some metadiscursive expressions appear in the native list only does not mean that they were not used at all by non-native students. It rather means that they were not statistically significant, or they did not meet the frequency and dispersion criteria set for this study. Apart from divergences in the bundle types used by L1 Italian and L1 English writers, a limited number of metadiscursive expressions were found to be shared between the two groups of students. The shared units were marked in bold in Table 1. The metadiscursive chunks occurring in both databases of essays, albeit with different frequencies, are *it is important to*, *it is possible to*, *in the case of* and *on the other hand*. Furthermore, as for their functional subcategorisation, data suggests a similar distribution of interactive and interactional metadiscursive bundles in the two corpora with the former outnumbering the latter. In particular, text-oriented bundles represent 58% and 64% of all identified four-word bundles in the Italian EFL and L1 English academic writing respectively. The slightly lower use of interactive metadiscursive constructions by L1 Italian vs. L1 English students is counterbalanced by a more significant recourse to interactional or participant-

oriented resources in the non-native (42%) vs. native corpus (36%). These preliminary quantitative findings set the basis for a closer qualitative analysis of similarities and differences in recurrent word combinations at a functional level.

Italian corpus: 4-word				BAWE: 4-word			
N.	bundles	Freq.	Texts	N.	bundles	Freq.	Texts
3	ON THE OTHER HAND	47	25	1	IT IS IMPORTANT TO	31	17
4	THAT IS TO SAY	47	15	2	IN THE CASE OF	29	15
5	IT IS POSSIBLE TO	39	17	3	AS A RESULT OF	25	13
7	I AM GOING TO	31	15	7	THE FACT THAT THE	21	18
11	IT IS IMPORTANT TO	29	20	9	CAN BE FOUND IN	18	8
13	AT THE SAME TIME	27	20	10	ON THE OTHER HAND	18	14
15	IN THE CASE OF	25	14	14	AN EXAMPLE OF THIS	16	8
21	I WOULD LIKE TO	23	11	15	FOR EXAMPLE IN THE	16	10
30	FOR THIS REASON THE	19	14	16	IN THE FORM OF	16	13
32	ON THE BASIS OF	19	13	18	IT WAS FOUND THAT	16	9
33	ONE OF THE MOST	19	15	21	CAN BE SEEN IN	15	11
42	THE POINT OF VIEW	17	10	22	FOR THE PURPOSES OF	15	11
44	IN THIS PAPER I	16	15	23	THAT THERE IS A	15	12
49	IN THIS CASE THE	15	14	24	HAVE BEEN FOUND TO	14	5
				29	IT IS POSSIBLE TO	13	7
				34	THAT THERE IS NO	12	12
				41	WITH RESPECT TO THE	11	5
				42	AS WELL AS THE	10	9
				44	IN TERMS OF THE	10	9
					IT HAS BEEN		
				45	SUGGESTED	10	7
				46	IT IS CLEAR THAT	10	10
TOTAL NUMBER OF 373 4-WORD BUNDLES				TOTAL NUMBER OF 341 4-WORD BUNDLES			

Tab. 1: Non-native and native students' essays: metadiscursive units with related raw frequency and the number of texts in which they occur

3.2 Qualitative analysis of metadiscursive bundles

The constructions shown in Table 1 were more thoroughly analysed in their linguistic context and their metadiscourse function. Specifically, according to Hyland's taxonomies of interactive *versus* interactional metadiscourse, or text-oriented *versus* participant-oriented bundles, the clusters discussed in quantitative terms in the previous subsection were examined in the metadiscursive role they play.

Notwithstanding some differences, the two corpora of L1 Italian and L1 English essays show a similar proportion of text-oriented or interactive bundles and participant-oriented or interactional expressions with text-oriented sequences ranking as the largest category in both corpora. Thus, it can be inferred that both native and non-native writers employ a larger number of metadiscursive constructions to frame discourse and guide the reader through the unfolding text in comparison with interactional metadiscourse to signal the presence of writers and their involvement with both content and readers.

By examining the list of metadiscursive expressions, what can be recognised are some elements of interactive metadiscourse which similarly occur in both sets of essays. More to the point, in an attempt to structure texts and organise arguments, both Italian EFL and L1 English students were found to prominently rely on expressions like *in the case of* and *on the other hand* (see Examples 1-4, my italics as in the other quotations).

- (1) The main problems in translating both idioms and fixed expressions can be summarized as follows: there may be no equivalent in the target language, especially *in the case of* culture-specific idiomatic and fixed expressions [...] (It. corpus)
- (2) Many researchers who collected actual data, however, found that in some cases, the exact opposite to Lakoff's claims were true, for example *in the case of* tag questions. Lakoff maintained that women used more tag questions than men as a syntactic device to express uncertainty and to seek assurance. (BAWE corpus)
- (3) A literal translation consists, more precisely, in borrowing items, calques and the literal translation itself; *on the other hand*, adaptation, equivalence, modulation and transposition are typical techniques of oblique translation. (It. corpus)
- (4) Prescriptivists, like Dr. Johnson, see the need for a known distinction between correct and incorrect forms of English. They would argue for inequality in varieties of English, saying that the prestigious SE and RP are better than non-standard English and regional dialects and accents. Descriptivism *on the other hand* takes a describing approach to the varieties of English and does not label them correct or incorrect. (BAWE corpus)

As illustrated by the examples above, *in the case of* serves as a signal to situate an argument or to frame it by limiting its conditions while *on the other hand* is used to establish contrastive links between the elements of the ongoing discourse. Apart from these shared bundles, Italian students' essays turn out to be significantly pervaded by transition and resultative constructions such as *that is to say*, *at the same time* and *for this reason the*. This is evidenced,

for instance, in the following Examples (5-7) in which these metadiscursive bundles manifest students' willingness to clarify, elaborate and add to a topic, as well as mark causative relations between the elements under discussion.

- (5) Clusters are long strings of words, *that is to say* groups of words which regularly follow each other throughout the text. By means of clusters, larger units of words can be identified, so that it becomes easier to find out which words belong together. (It. corpus)
- (6) In the translation from English into Italian it is essential to maintain the persuasive function of the text [...]. *At the same time* it is fundamental to reproduce the ST information by adapting the SL structures to the norms and conventions of the target language and culture. (It. corpus)
- (7) In translating promotional texts from English into Italian some problems can be found, for ex. the use of loan words, of direct and personal style, of short sentences with the reiteration of some words, etc. *For this reason the* translator needs to adjust the text to the target audience as it sounds natural to them. (It. corpus)

The extensive use of expressions by Italian students for topic elaboration and transition is accompanied by a less frequent recourse to metadiscursive constructions which contribute to framing and situating arguments by specifying cases and pointing to limitations. Some emblematic signposts pointing to this framing function can be seen in such clusters as *on the basis of*, *(from) the point of view* and *in this case the* (see Examples 8-10):

- (8) *On the basis of* this type of analysis, the meaning of the source text has to be then transferred to the target language with various strategies, depending mainly on text type and purpose. (It. corpus)
- (9) *From the point of view of* the lexis, the use of synonyms and omissions avoided an excessive recurrence of repetitions. (It. corpus)
- (10) These verbs are chosen because of the immediacy they convey and the purpose they serve: an invitation to perform something. *In this case, the* imperative form is used to the writer to intervene in the story and to give orders, to recommend something, or to give some suggestions. (It. corpus)

The results discussed so far seem to suggest Italian EFL students' preference for bundles that explicitly signal relations between stretches of discourse and, to a lesser degree, constructions that frame arguments and focus readers on a given case or some limitations. On the one hand, the repeated use of these transition constructions may come from a variety of sources including

emphasis by instructors on transition markers as key resources that “help create textual cohesion by signalling logical links between propositions” (Cao and Hu 2014, 19). On the other hand, the non-native students’ limited reliance on frame markers could point to their difficulties in developing argumentation and little experience in writing academic argumentative essays. The trend characterising Italian students’ compositions is counterbalanced by L1 English writers’ more pronounced preference for resultative and framing constructions over bundles of topic clarification and transition markers. Notably, apart from constructions such as *an example of this* (Example 11), *for example in the* and *as well as the* (Example 12), which establish relations of equivalence between the elements of an argument, more recurrent and diversified resultative and framing strings can be observed in the corpus of native writing. Some revealing instantiations of these two categories of clusters can be respectively recognised in forms such as *as a result of* (Example 13), *(due to) the fact that the* (Example 14), *it was found that, have been found to*, and *in the form of* (Example 15), *the fact that the, for the purpose of, with respect to the* (Example 16), and *in terms of the*.

- (11) This therefore indicated that males seem more confident in their speech and possibly use indefinite pronouns as a form of laziness, therefore enabling them to avoid detailed descriptions. *An example of this* is in the case of ‘but it’s like everything else,’ where the use of ‘everything’ encompasses the whole variety of events or situations (depending on the context of the sentence), enabling the male speaker to avoid describing each one individually. (BAWE corpus)
- (12) Another approach to the analysis of spoken discourse is interactional sociolinguistics, which focuses on looking at the speech features people use when interacting with each other. Turn-taking (as previously covered with conversation analysis) *as well as the* use of immediate responses and minimal responses are a large part of this approach. (BAWE corpus)
- (13) *As a result of* this examination of three types of instruction, each based on a different theory of language learning, I can now view my Persian learning in a more informed light. (BAWE corpus)
- (14) Therefore although the most people tend to respond in the same way the answers do not help us to determine how words are linked within the lexicon. The test in itself is considered to be unnatural *due to the fact that the* words are presented on their own. Normal speech would involve the stimuli being surrounded by other words therefore the process of retrieval would no doubt be different. (BAWE corpus)

- (15) The Rumelhart and McClelland (RM) 1986 PDP model differs from the dual mechanism approach in two major respects. Firstly, it claims that a single system, utilising phonological feature analysis of input items, can account for the acquisition, and adult formulation, of the past tense of both regular and irregular verbs, *in the form of* a neural network that directly connects input and output units. (BAWE corpus)
- (16) Lyons (1977:456) states that “No agreement has yet been reached on the semantic interpretation of expressions containing ‘any’.” and this comment appears to relate to both forms. Examples (2) and (13) illustrate the problem *with respect to the* unstressed form, or existential quantifier. (BAWE corpus)

Notwithstanding the overall centrality of interactive metadiscursive bundles in both native and non-native essays, the reiterated occurrence of resultative and framing constructions in L1 English writing could be interpreted as a manifestation of native students’ more mature writing skills and argumentative perspective being adopted in their essays. A variety of formulae were indeed used to “focus readers on a given case [...] or to emphasize aspects of an argument [...], or specify the conditions” (Pan, Reppen and Biber 2016, 68).

Furthermore, the prominence of text-oriented bundles is also foregrounded by the slightly minor role played by participant-oriented formulae or interactional metadiscursive devices. As for the clusters that convey interpersonal meanings, both native and non-native writing is interspersed with expressions of the writer’s commitment to the status of information, namely its certainty or uncertainty, and her/his evaluation of discourse elements. In this regard, *it is possible to* (Examples 17 and 18) and *it is important to* (Examples 19 and 20) were found in both corpora to express the author’s epistemic and attitudinal evaluation of what is discussed at centre stage.

- (17) By looking through concordances and keywords, *it is possible to* learn how metaphors are used to guarantee a successful conceptual communication in journalistic texts which deal with dynamic changes in a world-wide economic topic. (It. corpus)
- (18) From these examples, *it is possible to* infer that the pre-modifying noun seems to immediately precede the head and may also be pre-modified by an adjective. (BAWE corpus)
- (19) It is essential that the translator recognizes the function of idiomatic and fixed expressions in the source text to reproduce it; in particular *it is important to* recreate the same effect of the source text in the TL. (It. corpus)
- (20) First of all *it is important to* consider that there is no direct correlation between individual patterns and the meanings of the words which follow them. (BAWE corpus)

However, apart from the same resources exploited by native and non-native students to signal stance, the essays written by Italian EFL students were observed to be characterised by the author's personal projection into the text through the first-person pronoun *I*. In particular, *I am going to* (Example 21), *I would like to* (Example 22) and *in this paper I* (Example 23) appear to be clear expressions of the writer's intrusion into the text to organise the discourse and guide readers through it. In addition to personal self-projection into the discourse, the author's assessments of elements of discourse are marked by the use of the construction *one of the most* followed by evaluative adjectives such as *important* and *useful* (Example 24).

- (21) In this paper, *I am going to* translate the company presentation from the SL (English) to the TL (Italian), analysing it in terms of context, style, discourse and vocabulary. (It. corpus)
- (22) After analysing different linguistic means carrying marks of evaluation, *I would like to* draw some conclusions about the impact of evaluative language on the lexis used in fairy tales. (It. corpus)
- (23) *In this paper I* explained the meaning and the purpose of translation. I also examined the three main phases in the translation process namely comprehension of the ST, transfer of meaning and assessment of the target text. (It. corpus)
- (24) Among the large number of tools which linguists can rely on when studying language in certain fields, *one of the most important and useful* is the use of corpora. (It. corpus)

Explicit indicators of the author presence in statements of research purpose and outcomes turn out to be dispreferred by English students in favour of more impersonal constructions to express tentativeness and certainty as well as to pull the audience into the discourse. Corpus insights show that multi-word units such as *can be found in*, *can be seen in* (Example 25), *that there is a* (Example 26), *that there is no*, *it has been suggested* (Example 27) and *it is clear that* are strongly represented in L1 English writers' essays.

- (25) Similarly, there is the fact that past tense does not always exclusively refer to past time, as *can be seen in* the following example, 'If he came into the studio tomorrow,' whereby 'came' indicates the past tense but in the same construction you also have 'tomorrow,' which refers to something in the future. (BAWE corpus)
- (26) I think what can be seen through the research that has been done is *that there is a* clear link between class (incorporating status and prestige) with gender. (BAWE corpus)

- (27) Similarly women have been assigned the stereotype of asking lots of questions (Brouwer et al 1979 as in Coates 2004: 93), however *it has been suggested* that these too are used to maintain conversation (Maltz and Borker :213 whilst men's use of questions supports Tannen's (1991, Talbot 1998: 98) report talk. (BAWE corpus)

The evidence from the examples above substantiates that expressions of both tentativeness (e.g., *it has been suggested*) and assertion (*that there is a*, *that there is no* and *it is clear that*) are combined with engagement constructions which invite readers to enter the text and perform some actions (e.g., *can be found in* and *can be seen in*).

As for interactional metadiscursive bundles, the results of this study, which converge with previous research (Ädel and Erman 2012), show that native novice writers are more inclined than non-native authors to follow such academic writing conventions as impersonality and objectivity. While passive structures, the anticipatory *it* and existential *there*-constructions are distinctive of L1 English writing, Italian EFL students' essays are abundant with constructions using the first-person singular pronoun *I*. The greater personality in non-native compositions may be explained by EFL learners' lower level of academic writing experience and a lack of overt instruction in the use and frequency of use of both personal pronouns and more impersonal forms for the expression of stance and engagement in academic writing.

4. Discussion and conclusions

Results from this study further confirm general patterns found by research in the lexical bundle use by native and non-native writers, which were discussed in the Introduction to this paper.

In particular, despite the similar total number of metadiscursive constructions used by Italian EFL and L1 English students, native writers were found to produce more varied combinations than non-native ones. Furthermore, as for their functional characteristics, the analysis of Italian and English students' essays revealed a similar distribution of text-oriented and participant-oriented bundles with the former being more prominently resorted to than the latter by both groups of writers.

More interestingly, the examination of metadiscursive four-word formulae, which does not pretend to be exhaustive, suggests that native students are more mature writers in comparison with non-native ones. This is substantiated by L1 English authors' tendency to use more framing signals (e.g., *in the form of*, *with respect to the*, *in terms of the*), passive expressions (e.g., *can be found in* or *can be seen in*), existential *there* also in negative constructions (e.g., *that there is a* and *that there is no*), and *fact*-headed bundles (e.g., *the fact that the*). Moreover, the more sophisticated metadiscursive and argumentative nature of English writers' texts manifests

itself in the recourse to both tentative and assertive impersonal constructions which, apart from indicating better control of cautious language, suggest English students' preference for structures introduced by *it* that are used to disguise authorial interpretations and project a detached writing persona (e.g., *it has been suggested, it was found that, it is important to, it is clear that*). These elements prominent in native writing prove to be less statistically significant in non-natives' essays. Insights from the learner corpus show that Italian EFL students, whose level of proficiency in English is upper-intermediate/advanced, are more prone to using 'less sophisticated' interactive metadiscursive bundles and 'safer' phrases, or 'teddy bear phrases,' that they cling to and use repeatedly. These include markers of elaboration and transition (e.g., *that is to say, on the other hand* and *at the same time*), as well as self-projecting statements introduced by the *I*-form (e.g., *I am going to* and *I would like to*). The preponderant recourse to these metadiscursive clusters may arise from the interaction between two coexisting factors: the fact that Italian students are writing in English as a Foreign language and their lack of experience in academic writing and familiarity with academic discourse conventions.

In this light, despite the limited number of essays and four-word clusters selected and analysed, the study discussed in this paper could serve as a contribution to teaching the use of more diversified metadiscursive resources and academic discourse conventions to non-native students or at least to raising their awareness about the existing and extensive repertoire of devices available to academic writers. It can be hoped that further research—e.g., exploring a more extended range of metadiscursive bundles, increasing the number of essays under study and investigating the influence of L1 on EFL learners' use of metadiscourse—will contribute to a more complete examination of the different metadiscursive constructions produced by L1 Italian and L1 English university students in their essays.

Donatella Malavasi is Associate Professor of English Language and Translation at the University of Modena and Reggio Emilia (Italy). Her research interests lie in discourse analysis, corpus linguistics and business communication (*"The Multifaceted Nature of Banks' Annual Reports as Informative, Promotional and Corporate Communication Practices,"* 2010; *"Selling How Good We Are: An Analysis of Web-Based CSR Communication in 'Made in Italy' Companies,"* 2014). Her current research activity deals with the practices and strategies of knowledge dissemination across genres in business and academic contexts, and EDI communication (*"Building Trust through Corporate Identity: An Analysis of CSR Reports and Webpages,"* 2017; *"Academic Writing vs. Blogging: Paul Krugman as a Case Study,"* 2020;

“(Re-)framing diversity in discourse: Impact of recent social movements on corporate communication,” 2023).

Works cited

- Ädel, Annelie. *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins, 2006.
- Ädel, Annelie and Britt Erman. “Recurrent Word Combinations in Academic Writing by Native and Non-native Speakers of English: A Lexical Bundles Approach.” *English for Specific Purposes* 31.2 (2012): 81-92.
- Bamford, Julia and Marina Bondi, edited by. *Dialogue within Discourse Communities: Metadiscursive Perspectives on Academic Genres*. Tübingen: Niemeyer, 2005.
- Berkenkotter, Carol, Vijay K. Bhatia and Maurizio Gotti, edited by. *Insights into Academic Genres*. Bern: Peter Lang, 2012.
- Bhatia, Vijay K. *Analysing Genre: Language Use in Professional Settings*. London: Longman, 1993.
- Bhatia, Vijay K. *Worlds of Written Discourse: A Genre-Based View*. London: Continuum, 2004.
- Biber, Douglas and Federica Barbieri. “Lexical Bundles in University Spoken and Written Registers.” *English for Specific Purposes* 26.3 (2007): 263-286.
- Biber, Douglas, Susan Conrad and Viviana Cortes. “If you look at ...: Lexical Bundles in University Teaching and Textbooks.” *Applied Linguistics* 25.3 (2004): 371-405.
- Biber, Douglas, et al. *Longman Grammar of Spoken and Written English*. Harlow: Pearson, 1999.
- Cao, Feng and Guangwei Hu. “Interactive Metadiscourse in Research Articles: A Comparative Study of Paradigmatic and Disciplinary Influences.” *Journal of Pragmatics* 66 (2014): 15-31.
- Carrió-Pastor, María L., edited by. *Corpus Analysis in Different Genres: Academic Discourse and Learner Corpora*. London: Routledge, 2020.
- Chen, Yu-Hua and Paul Baker. “Lexical Bundles in L1 and L2 Academic Writing.” *Language Learning & Technology* 14.2 (2010): 30-49.
- Cortes, Viviana. “Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology.” *English for Specific Purposes* 23.4 (2004): 397-423.
- Crismore, Avon, Raija Markkanen and Margaret Steffensen. “Metadiscourse in Persuasive Writing: A Study of Texts Written by American and Finnish University Students.” *Written Communication* 10 (1993): 39-71.

- Dahl, Trine. "Textual Metadiscourse in Research Articles: A Marker of National Culture or of Academic Discipline?" *Journal of Pragmatics* 36.10 (2004): 1807-1825.
- Durrant, Philip. "Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories." *Applied Linguistics* 38.2 (2017): 165-193.
- Ellis, Nick C. "Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear." *Annual Review of Applied Linguistics* 32 (2012): 17-44.
- Esfandiari, Rajab and Fatima Barbary. "A Contrastive Corpus-driven Study of Lexical Bundles between English Writers and Persian Writers in Psychology Research Articles." *Journal of English for Academic Purposes* 29 (2017): 21-42.
- Flowerdew, John. *Academic Discourse*. London: Routledge, 2014.
- Granger, Sylviane and Fanny Meunier, edited by. *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins, 2008.
- Granger, Sylviane and Magali Paquot. "Lexical Verbs in Academic Discourse: A Corpus-driven Study of Learner Use." *Academic Writing: At the Interface of Corpus and Discourse*. Edited by Maggie Charles, Diane Pecorari and Susan Hunston. London: Continuum, 2009. 193-214.
- Güngör, Fatih and Hacer Hande Uysal. "A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars." *English Language Teaching* 9.6 (2016): 176-188.
- Habibie, Pejman and Ken Hyland, edited by. *Novice Writers and Scholarly Publication: Authors, Mentors, Gatekeepers*. London: Palgrave Macmillan, 2019.
- Hyland, Ken. "Academic Clusters: Text Patterning in Published and Postgraduate Writing." *International Journal of Applied Linguistics* 18.1 (2008a): 41-62.
- . *Academic Discourse: English in a Global Context*. London: Continuum, 2009.
- . "As Can be Seen: Lexical Bundles and Disciplinary Variation." *English for Specific Purposes* 27 (2008b): 4-21.
- . "Bundles in Academic Discourse." *Annual Review of Applied Linguistics* 32 (2012): 150-169.
- . *Metadiscourse*. London: Continuum, 2005.
- . "Metadiscourse: What Is It and Where Is It Going?" *Journal of Pragmatics* 113 (2017): 16-29.
- . "Participation in Publishing: The Demoralizing Discourse of Disadvantage." *Novice Writers and Scholarly Publication: Authors, Mentors, Gatekeepers*. Edited by Pejman Habibie and Ken Hyland. London: Palgrave Macmillan, 2019. 13-33.

- . "Persuasion and Context: The Pragmatics of Academic Metadiscourse." *Journal of Pragmatics* 30 (1998): 437-455.
- Hyland, Ken and Marina Bondi, edited by. *Academic Discourse Across Disciplines*. Bern: Peter Lang, 2006.
- Hyland, Ken and Carmen Sancho Guinda, edited by. *Stance and Voice in Written Academic Genres*. New York: Palgrave Macmillan, 2012.
- Hyland, Ken and Polly Tse. "Metadiscourse in Scholastic Writing: A Reappraisal." *Applied Linguistics* 25.2 (2004): 156-177.
- Kuteeva, Maria and Anna Mauranen. "Digital Academic Discourse: Texts and Contexts: Introduction." *Discourse, Context & Media* 24 (2018): 1-7.
- Li, Jie and Norbert Schmitt. "The Acquisition of Lexical Phrases in Academic Writing: A Longitudinal Case Study." *Journal of Second Language Writing* 18.2 (2009): 85-102.
- Mauranen, Anna. *Cultural Differences in Academic Rhetoric: A Textlinguistic Study*. Bern: Peter Lang, 1993.
- Mbodj, Ndeye Bineta and Scott A. Crossley. "Students' Use of Lexical Bundles: Exploring the Discipline and Writing Experience Interface." *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*. Edited by Ute Römer, Viviana Cortes and Eric Friginal. Amsterdam: John Benjamins, 2020. 115-133.
- Nam, Yunjung. "Give Constructions in Korean EFL Learner Writing." *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*. Edited by Ute Römer, Viviana Cortes and Eric Friginal. Amsterdam: John Benjamins, 2020. 33-58.
- Nesi, Hilary and Helen Basturkmen. "Lexical Bundles and Discourse Signaling in Academic Lecturers." *International Journal of Corpus Linguistics* 11.3 (2006): 283-304.
- Pan, Fan, Randi Reppen and Douglas Biber. "Comparing Patterns of L1 versus L2 English Academic Professionals: Lexical Bundles in Telecommunications Research Journals." *Journal of English for Academic Purposes* 21 (2016): 60-71.
- Paquot, Magali and Sylviane Granger. "Formulaic Language in Learner Corpora." *Annual Review of Applied Linguistics* 32 (2012): 130-149.
- Pérez-Llantada, Carmen. "Formulaic Language in L1 and L2 Expert Academic Writing: Shared and Distinct Usage." *Journal of English for Academic Purposes* 14 (2014): 84-94.
- Römer, Ute. "English in Academia: Does Nativeness Matter?" *Anglistik: International Journal of English Studies* 20.2 (2009): 89-100.

- Römer, Ute, Viviana Cortes and Eric Friginal, edited by. *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*. Amsterdam: John Benjamins, 2020.
- Scott, Mike. *WordSmith tools, 6.0*. [Online]. <http://www.lexically.net/wordsmith/version6>. 2012. Last visited 11/11/2023.
- Shaw, Philip. "Linking Adverbials in Student and Professional Writing in Literary Studies: What Makes Writing Mature." *Academic Writing: At the Interface of Corpus and Discourse*. Edited by Maggie Charles, Diane Pecorari and Susan Hunston. London: Continuum, 2009. 215-235.
- Simpson-Vlach, Rita and Nick C. Ellis. "An Academic Formulas List: New Methods in Phraseology Research." *Applied Linguistics* 31.4 (2010): 487-512.
- Sinclair, John M. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.
- . *Trust the Text: Language, Corpus and Discourse*. London: Routledge, 2004.
- Swales, John M. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press, 1990.
- . *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press, 2004.
- Vande Kopple, William J. "Some Exploratory Discourse on Metadiscourse." *College Composition and Communication* 26 (1985): 82-93.
- Wray, Alison. "Formulaic Sequences in Second Language Teaching: Principle and Practice." *Applied Linguistics* 21.4 (2000): 463-489.