

SENTENCE-INITIAL INTERACTIVE METADISOURSE AND TEXT COHERENCE IN EFL ARGUMENTATIVE ESSAYS¹

VIŠNJA PAVIČIĆ TAKAČ

Faculty of Humanities and Social Sciences

University of Osijek

e-mail: vpavicic@knjiga.ffos.hr

Abstract

In the field of L2 academic literacy an abundance of research on metadiscourse provides evidence of its overuse or underuse, particularly in comparison to native-speaker texts. These findings have important implications if efficient use of metadiscourse is considered indicative of text coherence. This study focuses on the interactive dimension of Hyland's (2005) metadiscourse model which encompasses the ways in which writers shape their relationship with the text and readers, i.e. how they relate, sequence and order propositions. Specifically, the study aims at identifying frequency and variation of L2 writers' use of sentence-initial metadiscursive elements (SI-ME), comparing these with native speakers' choices, and determining the relationship between SI-ME use and text cohesion and coherence. The results of SI-ME analyses, conducted on 80 essays written by Croatian EFL learners, show that L2 writers' use of SI-ME diverges from native speakers' in terms of range, density and accuracy, but is also idiosyncratic and creative, suggesting a strategic approach to constructing coherent texts.

Introduction

One of the most important aspects of academic literacy is writers' ability to use their knowledge of language elements to form articulate texts. Second language (L2) communicative competence models have included this ability under the notion of discourse competence (cf. Canale and Swain 1980; Canale 1983; Bachman 1990; Bachman and Palmer 1996; Celce-Murcia, Dörnyei and Thurrell 1995; Celce-Murcia 2007; Common European Framework of Reference (hereafter CEFR) 2001; Hymes 1972). Although there are noticeable differences in how these models define discourse competence and its relation to other communicative competence components, they all seem to share the idea that discourse competence includes the concepts of cohesion and coherence. Whereas the above listed models define cohesion in similar terms, there is no agreement on what coherence entails (Brown and Yule 1983; Canale 1983; Carell 1982; Dastjerdi and Talebinezhada 2006; Halliday and Hasan 1976; Savignon 1983; Van Dijk 1977; Widdowson 1978). Yet, most applied linguists concur that constructing a coherent text entails the ability to use metadiscourse, i.e. metadiscourse assists in establishing coherence (cf. Ädel 2006; Hyland 2005). Thus, appropriate usage of metadiscourse can be a useful indicator of coherence, which becomes of particular interest in the teaching and evaluation of second language academic writing. Empirical studies addressing the question of how metadiscourse works in achieving coherence, however, are still notably scarce. The present paper aims at exploring the relationship between the use of metadiscourse and coherence in argumentative essays written by users of English as a foreign language (EFL).

¹ This work has been supported by Croatian Science Foundation under the project IP-2016-06-5736.

Metadiscourse in L2 writing

Metadiscourse, as a relatively new term among linguists in current discourse analysis, has become widely used to denote "writing about writing, whatever does not refer to the subject matter being addressed" (Williams 1981: 211-212) or "text about text" (Ädel 2006: 2). Metadiscourse encompasses means of organising, interpreting, commenting or evaluating the content (i.e. propositions articulated in the text) with the purpose of helping the reader react to the text (Vande Kopple 1985). Metadiscursive elements are explicit textual devices writers employ to project themselves in their texts to interact with their receivers. Metadiscourse is "not a separate or separable set of stylistic devices that can either be included or not without affecting how a text is presented and read" (Hyland 2005: 23) and therefore makes a vital integrative feature of a text.

In the literature there is a variety of metadiscourse definitions, theories and taxonomies (cf. Ädel 2006; Crismore 1989; Crismore, Markkanen and Steffensen 1993; Hyland 2005; Thompson and Thetela 1995; Vande Kopple 1985; 2002). The existing models of metadiscourse have many similarities and some differences. Mostly, they follow Halliday's (e.g. 1994) systemic functional conceptualisation of language according to which language has three main functions: (1) interpersonal, referring to ways of using language to engage in interactions and express evaluations and personal feelings; (2) textual, referring to ways of using language to organise and structure the text; and (3) ideational, referring to the propositional content. In view of that, metadiscourse is seen as fulfilling either the interpersonal or textual functions, or both, and *not* fulfilling the ideational function. Generally, Ädel divides metadiscourse models into those taking the broad approach, what the majority of them do, and those taking the narrow approach, depending on whether they include stance markers, i.e. "linguistic material that expresses the speaker's attitude towards what is said" (Ädel 2006: 39) or completely exclude interpersonal elements, like Mauranen's (1993). Her own reflexive model, however, she positions as being between the broad and narrow approach. Ädel's (2006) model departs from the others by adopting as its basis Jakobson's (e.g. 1998) theory of communication functions of language (i.e. its three functions: the metalinguistic, the expressive and the directive) for which she claims to be better able to delineate bordering concepts, to include the dimensions of writer and reader, and, most importantly, to underscore reflexivity. Reflexivity is defined by Lyons (1977, as cited in Ädel 2006: 1) as "the capacity of natural language to refer to or describe itself". Its central position in the model gives the model its name and is reflected in Ädel's conceptualisation of metadiscourse as including "reflexive linguistic items that refer to the text itself as text or as language" and "linguistic items which reveal the writer's and reader's [...] presence in the text" (Ädel 2006: 2).

Among the above listed models, Hyland's (2005) interpersonal Model of Metadiscourse seems to fit adequately the present study for several reasons. First, it is constructed on the synthesis of previous major conceptions and taxonomies of metadiscourse – in particular Crismore, Markkanen and Steffensen's (1993), Thompson and Thetela's (1995) and Vande Kopple's (1985) – whereby taking into account their strong points, recognising their limitations and adding necessary modifications. The result is a more comprehensive categorisation scheme consisting of two main categories of metadiscursive elements interactive and interactional, the former being the focus of the present study. Second, it has been developed specifically for academic texts analysis which makes it applicable to the corpus of the texts analysed in this study. Next, it addresses texts written by non-native writers of English thus providing a structured basis for teaching academic writing. Finally, as Hyland's is one of the most widely

used models, the input provided by previous studies is beneficial for potential comparison of results across studies. What follows, therefore, is a brief presentation of the model based on Hyland's (2005) elaboration.

According to Hyland (2005), the significance of metadiscourse in writing is reflected in writer's need to make sure that the information they convey is understood and interpreted in accordance with their intention. Hence, writing (or for that matter speaking) is viewed as a social and communicative process between writers or speakers and readers or listeners. Successful writing presupposes an effective use of metadiscourse to establish and manage the social relations between writers and potential readers of the text. The two main categories of Hyland's model, the interactive and interactional, are thus the defining characteristics of any communication and they are expressed through a range of rhetorical features which perform more specific functions. The interactional metadiscourse concerns the writer's attempts to make his views explicit and involve the reader. It addresses ways writers conduct interaction by intruding and commenting on their message and includes the following categories: (1) hedges, i.e. means of withholding commitment and refraining from open dialogue (e.g. *might, perhaps*); (2) boosters, used to emphasize certainty or close dialogue (e.g. *definitely, surely*); (3) attitude markers, which express author's attitude towards propositions (e.g. *correctly, remarkably*); (4) self-mentions or explicit reference to the author (e.g. *I, the writer*); and (5) engagement markers, whose function is to explicitly build a relationship with reader (e.g. *follow, take as example*).

Interactive metadiscourse refers to writer's awareness of audience and the way he accommodates its probable knowledge, interests, rhetorical expectations and processing abilities. Its purpose is to shape and constrain a text to meet the reader's needs *and* to make sure the author's intended interpretations and goals are recovered. In other words, interactive metadiscourse guides the reader through the text. Thus, it refers to ways of organizing discourse, rather than experience. The interactive dimension contains five broad categories each of which fulfil a different function: (1) code glosses elaborate propositional meanings (e.g. *namely, for example., such as*), (2) endophoric markers refer to information in other parts of the text (e.g. *noted above, see fig. x*), (3) evidentials refer to information from other texts (e.g. *according to x, z states*), (4) frame markers refer to discourse acts, sequences or stages (e.g. *finally, to conclude*), and (5) transition markers express relations between main clauses (e.g. *in addition, but, thus, and*). In addition, the category of frame markers has four subcategories: (a) sequencing whose function is to sequence parts of text and to internally order an argument (e.g. *first, then*), (b) label stages that explicitly label text stages (e.g. *to summarize, in sum*), (c) announce goals (e.g. *I argue here, my purpose is*), and (d) shift topic (e.g. *let us turn to, well*).

As the description of Hyland's model suggests, it takes the broad approach because the term metadiscourse includes an expansive range of features and includes the interpersonal aspect, which, according to Ädel (2006) may create problems because the model becomes "too all-inclusive". Ädel's critique, although partially valid, will not be further discussed because it is primarily levelled against the interactional dimension of Hyland's model, which falls outside the scope of the present study.

Metadiscursive elements as linguistic features are not novel in L2 writing research but they have been studied under different labels such as for example discourse markers, cohesive devices, pragmatic markers, discourse or pragmatic particles, pragmatic expressions or connectives (cf. Jucker and Ziv 1998:1). This co-existence of different labels also signals a

variety of underlying theories and approaches which prevents drawing generalizable conclusions. However, what the majority of previous studies indicate is that L2 writers' use of those elements is characterized by over- or underuse, misplacement, errors and lack of variation (cf. Ädel 2006; Bagarić Medve and Čelebić 2012; Bagarić Medve and Pavičić Takač 2013; Bolton, Nelson, and Hung 2002; Chen 2006; Granger and Tyson 1996; Milton and Tsang 1993; Parise 2013; Pavičić Takač, Kružić and Vakanjac Ivezić in preparation). Evidently, there is a need for further, more comprehensive research anchored in solid theoretical models, what the present study intends to do.

The Present Study

This study has been conducted as part of a large scale research project entitled *Textual Coherence in Foreign Language Writing* (hereafter KohPiTekst) that deals with the description and cross-linguistic comparison of coherence features and the ways of developing coherence in texts written by writers in five different foreign languages (Croatian, German, English, French and Hungarian) and their respective first languages. One of its aims is to examine the relationship between the use of metadiscursive elements and the level of textual coherence. Inspired by previous related research and motivated by the lack of studies on the relationship between the patterns of metadiscourse use and coherence in EFL writing, the present study seeks to explore the use of sentence-initial interactive metadiscourse and its role in achieving coherence in academic argumentative essays written by EFL users. The use of interactive metadiscourse is compared against that of native speakers, but it must be pointed out that the native speaker texts do not serve as an idealised model of correctness, but as a reference point whose sole purpose is to provide "the basis for revealing the characteristics of learner language" (Johansson 1998: 7, as cited in Ädel 2006: 6).

Interactive metadiscourse is used to produce a text that the reader "is likely to find coherent and convincing" (Hyland 2005: 50). It plays an important role in producing and organising any text, but it features especially prominently in argumentative essays in which the writer is relying on metadiscursive markers to create relations between propositions, to sequence and refer to propositions in other text parts, as well as to establish a position and ensure that his or her ideas are followed and understood as intended (Ädel 2006; Crismore 1989; Granger and Tyson 1996). This, in addition to overall significance of argumentative essays as text type in modern society (Ädel 2006), justifies our selection of argumentative essay as the genre of writing to be analysed. Our focus on sentence-initial metadiscourse (hereafter SI-ME) has been prompted by several reasons. First, previous studies have documented an overuse of SI-ME in non-native writing. For example, Field and Yip (1992) and Granger and Tyson (1996) have found that non-native writers use significantly more sentence-initial connectors than NS writers which gives an impression of general overuse of metadiscourse. The second reason reflects the controversial status of metadiscourse in general, and especially in sentence-initial position in the Anglo-American writing style. As Parise (2013: 117) aptly states: "The rules of [SI-ME] use in writing have been subject to debate, myth and ambiguity." This is manifested in the different treatment SI-ME have received in different grammars where they have often been subjected to strict prescriptive rules regarding their placement (e.g.: *BABES* rule²), restriction to oral discourse and punctuation. Also, Anglo-American style guidebooks traditionally advise writers not to use metadiscourse excessively (cf. Bizup and Williams 2014). It may be speculated that native speakers of English receive more explicit instruction

² The *BABES* rule states that one should not begin a sentence with *but*, *and*, *because*, *especially* and *so*. See Bizup and Williams (2014) for a discussion.

in how metadiscourse affects style, whereas in the EFL context, due to writers' proficiency levels, more attention needs to be devoted to dealing with other language issues, such as lexical or grammatical. Metadiscursive elements have been differently categorized and labelled in reference books, which adds another point of confusion for any writer, let alone a non-native user of English who may largely rely on such resources for guidance. Our final reason for focusing on SI-ME stems from repeatedly shared impressions by our research project collaborators that participants "begin practically every sentence with a discourse marker".

The following were the research questions:

- (1) What is the frequency and variation of L2 writers' use of SI-ME?
- (2) How does L2 writers' use of SI-ME compare with native speakers' choices?
- (3) Is there a correlation between SI-ME use and text cohesion and coherence as well as overall text quality?

Methodology

Data collection procedure

The EFL corpus was compiled as part of the KohPiTekst research project. It consists of 80 texts written in English by EFL students, i.e. humanities and social sciences students at B2 proficiency level (CEFR 2001). Participants were asked to write a 200 – 230 words argumentative essay in which they were to present two points of view, express their opinion and give supporting arguments. The essays were hand-written during regular classes with no use of any external resources. In total, the corpus consisted of 19,657 tokens. The English native speaker corpus (hereafter: NS corpus) was compiled by selecting comparable argumentative essays from the Louvain Corpus of Native English Essays (LOCNESS; cf. Granger 1993). These essays were written by university students, native speakers of British or American English. The fact that there might be certain differences in the ways British and American writers use metadiscourse (cf. Ädel 2006) was disregarded because EFL learners in Croatia are exposed to both regional varieties in formal education and the media, and often coalesce their features in their language production. The NS corpus contained 31,062 tokens.

Data analysis procedure

For the quantitative analysis of SI-ME in the corpora, Sketch Engine (<http://www.sketchengine.eu>) was used. It is a multifunctional corpus manager and text analysis software with a tool for creating your own corpora (corpus architect). It allows for a case-sensitive search of words as well as multi-word phrases.

The analysis of corpora unfolded in a number of iterative stages. As has already been stated, the analysis was grounded in Hyland's (2005) model of metadiscourse. The starting point was his list of interactive metadiscourse items (Hyland 2005: 218-220). As all metadiscourse researchers are bound to agree, compiling a list of metadiscursive items is extremely challenging because their surface linguistic realisations may be remarkably diverse, ranging from morphemes and single lexical items to phrases, clauses and strings of sentences (Ädel 2006). Metadiscourse should be taken as "an open category to which writers are able to add new items according to the needs of the context" (Hyland 2005: 27). Therefore, no ready-made, all-inclusive and finite lists are possible. But, Hyland's list of items that may potentially realise metadiscourse functions offers a point of departure as well as of reference which may

jumpstart and facilitate the analysis. So, first the EFL corpus was searched for the items in the list. Next, all EFL texts were carefully read to identify other SI-ME not contained in Hyland's list. Finally, only those items identified in the EFL corpus were searched for in the NS corpus. This is why the method applied may be referred to, following Granger (1993), as partial (or one-way) contrastive interlanguage analysis. In addition to raw frequency, relative frequency (hits per 10,000 words) was calculated to permit comparison across corpora.

Since metadiscourse items are highly contextual, i.e. they may have different functions in different texts or fulfil two or more functions at the same time (cf. Ädel 2006; Hyland 2005), it was necessary to perform a fine-grained manual disambiguation of the SI-ME identified in the corpora. This analysis included deciding whether an item referred to the current text or a phenomenon external to the text, making a distinction between interactive or interactional meanings and categorising items according to Hyland's taxonomy. This required scrutinising the context in which an item was used, and/or running substitution tests. For example, in the EFL corpus the token *now* appeared six times in the initial position, but it was used as a time adverbial twice (example 1) and as a metadiscursive item in 4 instances, but with different functions. In two cases, *now* was used to label stages (example 2), but it was also used as a transition marker (example 3).

- (1) To my mind, city-life could become much better if the pollution would be taken more seriously. *Now* it doesn't seem as a big problem, but in the future it could become catastrophic.
- (2) *Now*, let us move on to the drawbacks.
- (3) Having more job opportunities, and therefore more choice, means that one can be more satisfied with a job they choose. *Now*, choice isn't only visible in the job department but also in the entertainment one.

An important issue to be dealt with prior to the analysis was the treatment of errors in EFL texts. As Thomas (2015: 12) puts it, learner-produced texts are typically “[...] spiced with deviant uses of language” but as any intervention in the text might compromise the authenticity and relevance of findings, it was decided to leave errors intact as written. Only spelling mistakes which would interfere with automatic computer searches were corrected (e.g. **Nevertheless* → *Nevertheless*).

Cohesion, coherence and overall text quality were evaluated by a group of trained raters within the KohPiTekst project who followed two sets of criteria. The first set included three elements: 1) text type (the degree to which a text complies with conventions of text structure in terms of introduction, body and conclusion; 20 points); (2) cohesion (which encompassed appropriate and efficient use as well as range of cohesive devices; 10 points) and (3) coherence (both local and global; 10 points). The second set included the following criteria: (1) task achievement (5 points), (2) vocabulary (range, appropriateness, orthography; 15 points), and (3) grammar (range and accuracy; 10 points). Each element was evaluated on the assessment scale ranging from 0 to 5, with a potential maximum of 70 points. The total number of points was taken as an indicator of overall text quality. For the purposes of this study points for the two cohesion subscales (maximum 10) and for the two coherence subscales (maximum 10) were taken as indicators of text cohesion and coherence.

Results

Frequency analysis of SI-ME

Tables 1 to 6 show the frequency of individual metadiscursive items in particular categories of SI-ME found in EFL and NS corpora. The items listed below the dotted line are additions to Hyland's list of metadiscursive items. In all tables, *f* stands for raw frequency, and *rf* for relative frequency (i.e. hits per 10,000 words).

Table 1: Frequency of SI code glosses

SI-ME: code glosses	EFL corpus		NS corpus	
	<i>f</i>	<i>rf</i>	<i>f</i>	<i>rf</i>
For (an) example	15 (14, 1)	7.66	4	1.29
For instance	4	2.04	2	0.64
That means (That's mean)	2 (1, 1)	1.02	-	-
One of <i>x</i>	4	2.04	7	2.25
For one	1	0.51	-	-
One example	1	0.51	-	-
To illustrate	1	0.51	-	-
Mentioning	1	0.51	-	-
Naming	1	0.51	-	-
To name	1	0.51	-	-
One thing	1	0.51	-	-

According to the results in Table 1, there were 11 types of sentence-initial code glosses in the EFL corpus. Out of a total of 32 tokens, *for example* (and its incorrect variant *for an example*) was used 15 times. In comparison, NS writers used SI-ME of this type far less frequently in their essays. The only exception is *one of x*, which appeared in the NS corpus slightly more frequently.

Table 2: Frequency of SI frame markers: sequencing

SI-ME frame markers: sequencing	EFL corpus		NS corpus	
	<i>f</i>	<i>rf</i>	<i>f</i>	<i>rf</i>
Firstly	31	15.69	9	2.89
Secondly	32	16.35	4	1.28
Another <i>x</i> /Other/The other	22 (20, 1, 1)	11.24	19 (-,-)	6.12
(The) first	6 (4, 2)	3.06	2 (-)	0.64
First of all	10	5.11	1	0.32
To begin with	7	3.58	-	-
Lastly	5	2.55	-	-
Finally	3	1.53	3	0.97
Second of all	3	1.53	-	-
Thirdly	3	1.53	-	-
First and foremost	2	1.02	-	-
Next	2	1.02	-	-
Subsequently	1	0.51	-	-
To begin	1	0.51	-	-

The last	1	0.51	-
Then	1	0.51	-
<hr/>			
At last	1	0.51	-
First of	1	0.51	-
First of to talk about	1	0.51	-
For start	1	0.51	-
For starter	1	0.51	-
Last but not least	1	0.51	-
Primarily	1	0.51	-
The last thing about	1	0.51	-

In the SI-ME category of sequencing (Table 2), there were four items clearly preferred by EFL writers: *firstly* and *secondly* (mostly used jointly as a pair), *another* or its variant, as well as *the first* and *first of all*. NS writers used some of the items found in the EFL corpus, but, again, far less frequently.

Table 3: Frequency of SI frame markers: label stages

SI-ME frame markers:	EFL corpus		NS corpus	
	<i>f</i>	<i>rf</i>	<i>f</i>	<i>rf</i>
All in all	14	7.08	1	0.32
To conclude	14	7.08	2	0.64
To sum (it) up	13 (11, 2)	6.64	1	0.32
In conclusion	4	2.04	2	0.64
Now	2	1.02	1	0.32
<hr/>				
In/At the end	3 (2, 1)	1.53	-	-
In retrospect	1	0.51	-	-
It can be concluded	1	0.51	-	-
One can conclude	1	0.51	-	-
To wrap up	1	0.51	-	-

To label stages, EFL writers seem to prefer those metadiscourse elements that signal the conclusion. Thus, *all in all*, *to conclude*, and *to sum (it) up* were used the most. These items occurred only once or twice in the NS corpus (Table 3).

Table 4: Frequency of SI frame markers: announce goals

SI-ME frame markers:	EFL corpus		NS corpus	
	<i>f</i>	<i>rf</i>	<i>f</i>	<i>rf</i>
In this essay	5	2.55	-	-
In this composition	1	0.51	-	-
In this writing assignment	1	0.51	-	-
<hr/>				
In order to (make a conclusion)	1	0.51	-	-
Today	1	0.51	-	-

As Table 4 shows, in the EFL corpus there were only five different items used to announce goals. The phrase *in this essay* was used five times and there were two more examples of its synonymous variations. None of these SI-ME were found in the NS corpus.

The next subcategory of frame markers includes items whose function is to shift topic but only six tokens were found in the EFL corpus. Only two appeared in the NS corpus (Table 5).

Table 5: Frequency of SI frame markers: shift topic

SI-ME frame markers: shift topic	EFL corpus		NS corpus	
	<i>f</i>	<i>rf</i>	<i>f</i>	<i>rf</i>
Now	2	1.02	1	0.32
As for	1	0.51	-	
Speaking of	1	0.51	-	
When it comes to	1	0.51	1	0.32
When we talk about	1	0.51	-	

The category of transition markers (Table 6) is by far the most numerous one in terms of both types and tokens. The majority of SI-ME that mark a transition in the text were used far more frequently in the EFL corpus. The most popular phrase was *on the other hand*, which was, interestingly enough, not always used in combination with *on (the) one hand*, at least in the sentence-initial position. Also very frequently used were the following *but*, *also* and *furthermore*. In two cases *but* was immediately followed by *on the other hand*. The NS corpus differs in this respect: *on the other hand* was found only twice, *but* and *also* were used much more rarely, and *furthermore* was not used at all. In this category there were four SI-ME whose relative frequency was higher in the NS corpus: *however*, *therefore*, *after all* and *despite*. The use of *so* was very similar in both corpora.

Table 6: Frequency of SI transition markers

SI-ME transition markers	EFL corpus		NS corpus	
	<i>f</i>	<i>rf</i>	<i>f</i>	<i>rf</i>
On the other hand	48	24.522		0.64
But	36	18.3911		3.54
Also	27	13.7918		5.79
Furthermore	22	11.24-		
However	13	6.64	48	15.45
Moreover	13	6.64	1	0.32
And	9	4.59	2	0.64
On (the) one hand	9 (5, 4)	4.59	1	0.32
While	8	4.09	4	1.29
On/At the other side	7 (6, 1)	3.58	-	
Although/Even though	7 (5, 2)	3.58	9 (8, 1)	2.89
So	7	3.58	11	3.54
Nevertheless	4	2.04	-	
Therefore	4	2.04	10	3.22
Because	3	1.53	1	0.32
In addition	3	1.53	-	

On the contrary/contrary to this	3 (2,1)	1.53	2	0.64
Thus	2	1.02	2	0.64
Besides	1	0.51	1	0.32
Further	1	0.51	-	
Since	1	0.51	-	
Still	1	0.51	-	
Then again	1	0.51	1	0.32
Whereas	1	0.51	-	
What is more	4	2.04	-	
Not to mention	3	1.53	-	
Now	2	1.02	-	
As well as (that)	2	1.02	2	0.64
After all	1	0.51	2	0.64
Despite	1	0.51	5	1.61
Good things aside	1	0.51	-	
(If we) look from one point of view	1	0.51	-	
Looking from another point of view	1	0.51	-	
Opposed to	1	0.51	-	
Parallely (→ "in contrast")	1	0.51	-	
That being said	1	0.51	-	

Table 7 summarises the results of the frequency analysis. Both raw frequency and relative frequency analysis point to a higher use of SI-ME by EFL writers in all categories of SI-ME. The TTR value showing the ratio between the total number of SI-ME and different types of SI-ME was calculated only for the EFL corpus. It indicates a relatively low lexical variation. No instances of SI endophoric markers or evidentials were found in the EFL corpus.

Table 7: Summary of SI-ME frequency analysis

SI-ME	EFL corpus			<i>rf</i>	NS corpus		
	<i>f</i> <i>tokens</i>	<i>types</i>	<i>TTR</i>		<i>f</i> <i>tokens</i>	<i>types</i>	<i>rf</i>
Code glosses	32	11	0.34	16.33	13	3	4.18
Endophoric markers	-						
Evidentials	-						
Frame markers: total	207	44	0.21				
FM sequencing	138	24	0.17	70.33	38	6	12.22
FM label stages	54	10	0.18	27.43	7	5	2.24
FM announce goals	9	5	0.55	4.59	-		
FM shift topic	6	5	0.83	3.06	2	2	0.64
Transition markers	240	36	0.15	122.59	133	19	42.77
TOTAL	685	134	0.20		193	35	

Correlation analysis

Table 8 shows the minimum, maximum as well as mean and standard deviation of frequency of SI-ME use by EFL writers. On average, EFL writers used 6.3 (SD = 2.5) SI-ME per text and they used 5.8 different metadiscursive items per text. EFL writers achieved a relatively high average score on both cohesion and coherence (3.9). As for overall text quality, participants scored 55.4 points on average with scores ranging from 28 to the maximum of 70 points. As for the relationship between SI-ME use and cohesion, coherence and text quality, there was a significant positive correlation between cohesion and SI-ME types and tokens, as well as between coherence and SI-ME tokens. As for overall text quality, it is positively correlated with both SI-ME tokens and SI-ME types.

Table 8: Means, standard deviations and Pearson correlation coefficients for the relation of SI-ME and cohesion, coherence and overall text quality

	min	max	M	SD	1	2	3
Cohesion	1.5	5	3.9	.89			
Coherence	1	5	3.9	.94			
Overall text quality	28	70	55.4	9.6			
SI-ME tokens	1	13	6.3	2.5	.291**	.227**	.320**
SI-ME types	1	13	5.8	2.3	.308**	.201	.327**

**Correlation is significant at the 0.01 level (2-tailed)

5. Discussion

By looking at the results of the frequency analysis (tables 1 to 6, and especially 7) one would come to two principal conclusions. The first conclusion is that, much in line with previous studies, EFL writers tend to overuse SI-ME. This is suggested by the comparison between the EFL and NS corpora which exposed great discrepancies in raw and relative frequencies of practically all SI-ME. The results also corroborate high frequency of a particular type of metadiscursive elements, namely the transition markers (cf. Hyland 2005; Shi and Han 2014). This is not the least unexpected, because transitions are an important characteristic of argumentative essays for they indicate internal connections and relationships between ideas in the text.

The frequency analysis also showed that one of the characteristics of EFL writers' use of SI-ME is their preference of particular individual items. EFL writers seem to have adopted specific items that they keep using to express certain functions. This may be interpreted as a consequence of exposure and explicit (over)teaching of lists of metadiscourse elements from which EFL writers pick and choose – in what appears to be an indiscriminate manner – those that they feel confident using (such as *but*), those that may appear more sophisticated (such as *furthermore*), or those that may serve the purpose of completing the task requirements (such as *firstly* and *secondly* to list arguments, or *on the other hand* to introduce a counterargument or even to reach the required number of words). As Ädel (2006: 19) states, "the implied reader exerts considerable influence on the writer and, consequently, on the text". Translated to writing argumentative essays in test-taking situations, such was the one in which our participants found themselves, the implied reader is the evaluator, and writers are well-aware of both the pending evaluation and the criteria used to judge their writing. What is more, they had been trained to meet them. One should not forget that high frequency of metadiscursive

elements in an essay may result from writers' targeting the required number of words (cf. Ädel 2006: 109). Thus, EFL writers' choices of particular metadiscursive items, overall high frequency of SI-ME use, as well as their cognizance of the evaluation process can be taken as indicators of the washback effect. Although there was some variation in the selection of SI-ME by EFL writers, it was relatively low (TTR = 0.20). Many items were found only once or twice in the EFL corpus and these examples of *hapax legomena* found in the EFL corpus were extremely rare or non-existent in the NS corpus.

However, these findings, although supported by numerical data, should not be taken at face value because they are an unequivocal effect of the methodology applied in the study. Since the starting corpus was the EFL corpus, only those items found in that corpus were searched for in the NS corpus. Thus, the results pointing to the overuse of SI-ME may be attributed to NS' *underuse* of those particular items. NS may use SI-ME equally frequently, but may opt for different linguistic realisations. It may also be speculated that NS in fact refrain from using metadiscourse in sentence-initial position because they adhere to prescriptive rules briefly mentioned above that they have probably been taught in their formal education. This issue remains to be explored in one of the future studies.

There was another interesting, but not surprising finding that further explains the differences between EFL and NS use of SI-ME in the present study. Namely, many examples of SI-ME uses in the EFL corpus, especially of the *hapax legomena*, were actually erroneous and therefore would not appear in the NS corpus. Although error analysis was beyond the scope of the present investigation, the process of categorisation of SI-ME required a closer inspection of those items. This revealed that the majority of errors were of formal nature pointing to problems with writers' language proficiency (examples 4 and 5):

- (4) *That's mean* also that air is polluted more than in an areas that are less suburb.
- (5) *At last*, crime rate in cities is getting bigger every day.

But, some of the errors point towards communication strategies that are employed by EFL writers with the purpose of reaching a communicative goal while compensating for the gap in their linguistic repertoire. Thus, examples 5, 6, 7 and 10 might illustrate the use of approximation, i.e. using an alternative item which expresses the meaning of the target term as closely as possible, whereas examples 8 and 9 indicate an attempt at coining expressions by applying L2 morphological rules, or by translating a similar expression from L1. Although communication strategies are used with the intention of promoting communication, some may in fact result in an inadequate or imprecise expression leading to communication breakdown or misunderstanding (see example 9).

- (6) *Opposed to* living in a village, city life is hectic and everyone is always late.
- (7) *On the other side* living in the city can sometimes be outrageously overpriced,
- (8) *Not to mention* schools and kindergartens.
- (9) *Parallelly*, villages are nicer places because of the quiet.
- (10) *In retrospect* the city can also carry some bad connotations

Another pronounced feature of EFL SI metadiscourse is the use of stylistically inappropriate items. The examples 11 to 15 illustrate the use of SI-ME that would conventionally be used in oral communication and allude to EFL writers' lack of awareness of stylistic constraints (cf. Granger and Tyson 1996).

- (11) *For start* let's talk about the pros of living in a city.

- (12) *For starter* it has bigger more efficient network of public transportation.
- (13) *To wrap up* with the advantages, let's say something about spending one's free time.
- (14) *Speaking of* the traffic, traffic jams are something one should prepare for
- (15) *When we talk about* the disadvantages of living in the city we can also connect it to the advantage of living in the city.

The correlation analysis showed that there is a relationship between SI-ME and text coherence, cohesion and overall text quality: the more SI-ME EFL writers use, the higher levels of cohesion, coherence and text quality they achieve. Whether they vary them seems to make a difference only when it comes to coherence (see Table 8). Although there has been some evidence of the positive correlation between text quality and metadiscourse (cf. Shi and Han 2014), these results stand in sharp contrast with the findings of the study we previously carried out on the same EFL corpus (cf. Pavičić Takač and Vakanjac Ivezić 2018) that showed no statistically significant correlations between three subcategories of frame markers (sequencing, labelling stages and shifting topic) and cohesion and coherence. The only significant correlation was found between cohesion and frame markers of announcing goals, but it was a negative one ($p = -.237$). Nonetheless, the findings of the present study show that SI-ME seem to play a positive role in establishing cohesion and coherence. To what can this be attributed? It may be hypothesised that sentence-initial position renders items more noticeable which makes L2 writers aware of what type of meaning they should formulate in relation to other propositions. For example, an introductory *on the other hand* induces a counterargument. Also, EFL writers frequently opt for pairs of metadiscursive items: if they use *firstly*, they also use *secondly* which makes them think of additional arguments they may list. As a result, the produced text is more coherent. Naturally, for this to work, EFL writers must know what an item exactly means, for a misuse of an item may lead to misunderstanding or coherence disruption, as illustrated in example 16 where *nevertheless* does not signal the denial of the previous proposition as it should.

- (16) Moreover, you can meet a lot of interesting people, for example chances of meeting their significant other or perhaps a business partner are much higher. *Nevertheless*, something is always happening, there is always a party somewhere, ...

The saliency of sentence-initial items may have the same effect on evaluators who then identify texts abundant in such items as more coherent. Thus, relatively highly assessed cohesion and coherence may be an effect of the frequent use of SI-ME.

Conclusion

The present study addressed the relationship between sentence-initial metadiscourse and coherence in argumentative essays written by EFL writers. The analysis revealed a number of noteworthy findings. First, compared to NS writers, EFL writers tend to use a particular set of SI-ME considerably more often but the variability of linguistic realisations they choose to express particular metadiscursive functions is limited. Although this result points to a general trend in non-native use of metadiscourse, it may come as a consequence of NS writers using a different set of SI-ME or attempting to reduce metadiscourse use because they might have been taught to do so for stylistic reasons. Second, the EFL corpus evidence shows that non-native writers mix registers using both very formal and informal colloquial expressions in their writing. Next, the use of SI-ME by EFL writers is occasionally erroneous or imprecise. These findings suggest that EFL writers' use of SI-ME is a consequence of exposure and

explicit teaching of, most probably, lists of expressions whose meanings and functions are not elaborated, contrasted or studied contextually. For example, items with similar functions, such as that of contrast, are presented as synonyms which leads to their indiscriminate use. Also, EFL learners lack of awareness of nuances of meaning or insecurity as to what they might be may motivate them to keep using the same items, which are usually those that are prototypical, cover a wider range of meanings, have fewer constraints or have a direct equivalent in their L1 (such as *but*). A close examination of the items found in the EFL corpus suggested that EFL writers' use of SI-ME is strategic which refers to both their employment of communication strategies to bridge the gap between their linguistic knowledge and communication needs and finding ways of complying with task requirements. Thus, when faced with problems in communicating the intended meaning, they use invented items they coined either by following a model in English (e.g. *on the other side* instead of *on the other hand*) or attempting to translate a similar item from their L1 (e.g. *not to mention*). Although in most cases the newly-coined items fulfil the intended metadiscursive function, occasionally they are a complete miss, such as the use of *parallelly* to mean *in contrast*. As for the effect of task requirements, EFL writers are aware that the use of metadiscursive elements is part of the evaluation criteria and may calculatingly use them to meet this criterion. Also, by adding metadiscourse items, one comes closer to the required number of words without having to think about propositions and elaborations.

Finally, our findings suggest that SI-ME may contribute to a more cohesive and coherent text. A possible explanation is that it is their position at sentence beginnings that guides writers to conceptualise and formulate the propositions that follow SI-ME in an adequate and logical manner, thus creating coherent and cohesive texts. But, the significant correlation between SI-ME and text coherence and cohesion may be an effect of the way cohesion and coherence are evaluated: evaluators might also perceive and, consequently, rate texts in which SI-ME are prominently featured as more coherent.

The findings of this study seem to echo the implications for teaching of metadiscourse that have been voiced in the past (cf. Davatgari Asl and Moradinejad 2016; Granger and Tyson 1996; Milton and Tsang 1993; Sadeghi and Kargar 2014). To reiterate, explicit teaching of metadiscourse must go beyond presenting simplistic lists of what seem to be mutually interchangeable linguistic items. The approach that can be highly recommended is data-driven learning because it puts emphasis on meaning *in context* thus allowing for disambiguation of polysemous, ostensibly synonymous or multifunctional items, as well as a more fine-grained analysis of examples of typical academic moves. A focus on developing learners' sensitivity to register is also called for. Last but not least, just as teaching of writing should include exploration of all means of expressing metadiscursive functions, it should include discussions of ways of establishing cohesion and coherence other than by employing explicit metadiscursive items.

The present study has brought forward interesting features of metadiscourse patterns in EFL academic writing, but even more exciting are the avenues of future research these findings open up. Future research might apply a mixed-method approach to analysis of coherence and cohesion (e.g. thematic progression/topical structure analysis) which would mitigate the effect of evaluators' perception by generating more objective indicators. Next, conducting a more detailed and comprehensive study of metadiscourse and its role in establishing cohesion and coherence would yield more valid, reliable and generalisable conclusions. A final intriguing question that warrants further exploration is the role of transfer of linguistic and cultural norms from native or other foreign languages in creating academic texts. Our KohPiTekst

project, we believe, provides a platform for conducting more systematic inquiries of at least some of these issues.

REFERENCES

- Ädel, A. 2006. *Metadiscourse in L1 and L2 English*. Amsterdam/Philadelphia: John Benjamins.
- Bachman L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford etc.: Oxford University Press.
- Bachman, L. & A. Palmer. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford etc.: OUP.
- Bagarić Medve, V. & I. Čelebić. 2012. Bewertung der Diskurskompetenz: holistischer vs. analytischer Ansatz. In L. Pon, V. Karablić & S. Cimer (Eds.). *Applied Linguistics Today: Research and Perspectives – Angewandte Linguistik heute: Forschung und Perspektiven* (pp 33-49). Frankfurt a/M etc.: Peter Lang.
- Bagarić Medve, V. & V. Pavičić Takač. 2013. The influence of cohesion and coherence on text quality: A cross-linguistic study of foreign language learners' written production. In E. Piechurska-Kuciel & E. Szymanska-Czaplak (Eds.). *Language in Cognition and Affect* (pp. 111-131). Berlin-Heidelberg: Springer.
- Bizup, J. & J. Williams. 2014. *Style: Lessons in Clarity and Grace* (11th edition). Harlow: Pearson.
- Bolton, K., G. Nelson & J. Hung. 2002. A corpus-based study of connectors in student writing: Research from The International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics* 7,2: 165-182.
- Brown, G. & G. Yule. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Canale, M. 1983. From communicative competence to communicative language pedagogy. In J.C. Richards & R.W. Schmidt (Eds.). *Language and Communication* (pp. 2-27). London: Longman.
- Canale, M. & M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1,1: 1-47.
- Carrell, P. 1982. Cohesion is not coherence. *TESOL Quarterly* 16: 479-488.
- Celce-Murcia, M. 2007. Rethinking the role of communicative competence in language teaching. In E. Alcón Soler & M.P. Safont Jordà (Eds.). *Intercultural Language Use and Language Learning* (pp. 41-57). Dordrecht: Springer.
- Celce-Murcia, M., Z. Dörnyei & S. Thurrell. 1995. Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6,2: 5-35.
- Chen, Ch. W. 2006. The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics* 11,1: 113–130.
- Council of Europe. 2001. *Common European Framework of Reference: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Crismore A. 1989. *Talking with Readers: Metadiscourse as Rhetorical Act*. New York: Peter Lang
- Crismore, A., R. Markkanen & M. Steffensen. 1993. Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students. *Written Communication* 10,1: 39-71.
- Dastjerdi, H. V. & R. M. Talebinezhad. 2006. Chain-preserving deletion procedure in cloze: a discursal perspective". *Language Testing* 23,1: 58-72.

- Davatgari Asl, H. & A. Moradinejad. 2016. The Effect of Explicit Instruction of Discourse Markers on Iranian EFL Learners' Speaking Ability. *Journal of Applied Linguistics and Language Research* 3,5; 190-202.
- Field, Y. & L. Yip. 1992. A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal* 23, 1: 15-28.
- Granger, S. 1993. The International Corpus of Learner English. In J. Aarts, P. De Haan and N. Oostdijk (Eds.). *English Language Corpora: Design, Analysis and Exploitation* (pp. 57-71). Amsterdam and Atlanta: Rodopi.
- Granger, S. & S. Tyson. 1996. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes* 15,1: 17-27.
- Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar* (2nd edition). London: Edward Arnold.
- Halliday, M. A. K. & R. Hasan. 1976. *Cohesion in English*. London: Longman Pub Group.
- Hymes, D. 1972. On communicative competence. In J.B. Pride & J. Holmes (Eds.). *Sociolinguistics: Selected readings* (pp. 269-293). Penguin, Harmondsworth.
- Jakobson, R. 1998. On Language: Roman Jakobson. L. R. Waugh & M. Monville-Burston (Eds.). Cambridge, MA and London: Harvard University Press.
- Jucker, A. H. & Y. Ziv. 1998. Discourse markers: Introduction. In A. H. Jucker & Y. Ziv (Eds.). *Discourse markers. Descriptions and Theory* (pp. 1-13). Amsterdam and Philadelphia: John Benjamins.
- Mauranen, A. 1993. *Cultural Differences in Academic Rhetoric: A Textlinguistic Study*. Frankfurt am Main: Peter Lang.
- Milton, J. C. & E. S. C. Tsang. 1993. A corpus-based study of logical connectors in EFL students' writing: directions for future research. In R. Pemberton & E.S.C. Tsang (Eds.). *Studies in Lexis* (pp. 215-246). Hong Kong: The Hong Kong University of Science and Technology Language Center.
- Pavičić Takač, V., B. Kružić & S. Vakanjac Ivezić. (in preparation) A corpus-driven exploration of discourse markers in L2 academic texts. Proceedings of the 1st International Conference on Corpus Analysis in Academic Discourse. Valencia, Spain.
- Pavičić Takač, V. & S. Vakanjac Ivezić. 2018. Frame markers and coherence in L2 argumentative essays. Paper presented at the University of Zagreb Round Table: Empirical Studies in Applied Linguistics. Zagreb, Croatia.
- Parise, P. 2013. Mind the Gap: Interpretations of Sentence-Initial Coordinating Conjunctions in Linguistic Studies of L2 and L1 Writing. *BULLETIN of Kanagawa Prefectural Institute of Language and Culture Studies* 2: 117-130.
- Sadeghi, B & A. Kargar. 2014. The Effect of Explicit Instruction of Discourse Markers on EFL Learners' Writing Ability. *International Journal of Educational Investigations* 1,1: 328-338.
- Savignon, S. J. 1983. *Communicative Competence: Theory and Classroom Practice. Texts and Contexts in Second Language Learning*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Shi, W. & J. Han. 2014. Research on Writing Samples from the Perspective of Metadiscourse. *English Language Teaching* 7, 11: 151-158.
- Sketch Engine. Available at <http://www.sketchengine.eu>
- Thomas, J. 2015. *Discovering English with Sketch Engine*. Brno: Laptop Languages.
- Thompson, G. & P. Thetela. 1995. The sound of one hand clapping: the management of interaction in written discourse. *TEXT* 15,1: 103-27.
- Vande Kopple, W. J. 1985. Some Exploratory Discourse on Metadiscourse. *College Composition and Communication* 26: 82-93.

- Vande Kopple, W. J. 2002. Metadiscourse, discourse, and issues in composition and rhetoric. In E. Barton & G. Stygall (Eds.). *Discourse Studies in Composition* (Vol. 91-113). Cresskill, NJ: Hampton Press.
- van Dijk, T. A. 1977. *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*. New York: Longman.
- Widdowson, H. G. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.
- Williams, J. W. 1981. *Style: Ten Lessons in Clarity and Grace*. Glenview, IL: Scott, Foresman.