

Spoken Discourse Markers in Learner Academic Writing

Alė Šimčikaitė

crossref <http://dx.doi.org/10.5755/j01.sal.0.20.1196>

Abstract. The present article focuses on the use of spoken features in learner academic writing. It aims to analyze the spoken-like nature of learner academic writing through the use of informal or semi-formal discourse markers in their academic essays. Non-experimental methods of data collection were chosen to achieve these objectives (the data was collected from language corpora): quantitative method was used (the frequency counts of discourse markers were indicated); qualitative and contrastive methods were used (the types of discourse markers were discussed and the comparative analysis between the three corpora was made). The results of the present investigation revealed that both the Lithuanian learners and the native learners use stylistically inappropriate (more typical of speech than of academic writing) discourse markers in their academic essays. In contrast to the native learners, the Lithuanian learners tend to use more of spoken discourse markers in their essays. Fifteen functional categories of the spoken discourse markers were distinguished. They helped to disclose which functional types of discourse markers tend to appear more often in the Lithuanian learners' essays. Other spoken features were briefly observed during the analysis of the discourse markers too. Both spoken discourse markers and other lexical items more typical of speech than of academic writing contribute to the overly oral tone of the learners' academic essays.

Key words: *spoken features, written/spoken English, discourse marker, academic writing, functional categories.*

Introductory Observations

Until recently much more attention has been paid to the grammar of written English, while the grammar of spoken English was not clearly described. It is the main reason why differences between spoken and written English still cause some confusion for learners. In recent years, researchers have become more interested in the features of spoken language, therefore, the two forms of language were opposed and different features of written and spoken English were distinguished. Scholars like Leech and Svartvik (1994), Biber et al. (1999), Pridham (2001), Huddleston and Pullum (2002), Carter and McCarthy (2006) indicate that spoken and written English differ in terms of grammar, vocabulary, formality and spontaneity. Colloquial speech, as stated by Biber et al. (1999, p.1121), has *vernacular* grammar and often includes structures that are not appropriate to standard written English. For example: "the multiple negative construction (e.g. *Don't say I never gave you nothing*) and the double comparative construction" (e.g. *Sometimes, that is so, so much more easier to follow*). Carter and McCarthy (2006, p.167) also mention split infinitives (e.g. *We decided to immediately sell it*), singular nouns after plural measurement expressions (e.g. *He's about six foot tall*) etc. Academic writing is one of the most difficult skills to master that has quite complex structures and is more formal and impersonal in style than everyday language. However, it has been noticed that some of the features of spoken English which contribute to the oral tone of a written work tend to appear in learner academic writing. The appearance of the computer corpus in the early 1960s gave rise to a branch of linguistics called *corpus linguistics*. This new discipline provides the opportunity to uncover new facts about language and to explore cross-linguistic differences. Scholars like Kennedy (1998), Granger (1998), Granger et al. (2002), Meyer (2002), Halliday et al. (2004) and Baker (2006) were

interested in corpus linguistics. Granger et al (2002, p.1) states that

it is neither a new branch of linguistics nor a new theory of language, but the very nature of the evidence it uses makes it a particularly powerful methodology, one which has the potential to change perspectives on language.

According to Granger (1998, p.3), the emergence of a corpus of advanced learner English has made it possible to come up with a new, more concrete approach to the lexical features of learner language. Corpora have been designed to answer questions at various linguistic levels on the prosody, lexis, grammar, discourse patterns, pragmatics etc. Discourse markers- words "that have little lexical meaning and appear on the periphery of clause structure" (Masaitienė, 2003, p.66) serve a very important role in structuring the discourse. These lexical items that help to link segments of the text and make it coherent, is the subject matter of this paper. A lot of attention has been paid to discourse markers by various scholars, however, little attention has been paid on the stylistic peculiarities of these lexical items. Discourse markers are sensitive to discourse type. Therefore, informal or semi-formal discourse markers used in an academic essay might contribute to the overly oral tone of the whole essay. The fact that different discourse markers can be used in written and spoken English and the possibility to use a computer based corpora determined the aims of this research. Therefore, the aims of the present paper are manifold: to find out whether advanced Lithuanian learners of English use stylistically inappropriate discourse markers to the same extent as native (British and American) learners; to disclose which functional categories of the spoken discourse markers prevail in their academic essays as well as to verify whether the learners tend to use other spoken features in their academic essays. To achieve these aims non-experimental methods of data collection were used.

The data for the present research come from *LICLE* (Lithuanian sub-corpus of International Corpus of Learner English). It consists of the academic essays produced by advanced Lithuanian learners of the English language. Two types of essay writing (argumentative essays and literature examination papers) were written by advanced students of English Philology. *LICLE* contains 154 992 words of academic writing. The native speakers' data is stored in the British, and American segments of *LOCNESS* (Louvain Corpus of Native English Essays). *LOCNESS-BR* contains 111 127 words and *LOCNESS-US* contains 168 231 words of academic writing. Quantitative and qualitative methods were used too. The frequency of discourse markers in learner academic writing was indicated. In terms of qualitative methods, it was important to analyze the types of discourse markers. Apart from this, the contrastive method was applied. The comparative analysis was made between the Lithuanian learner corpus data and the written native (British and American) corpus data.

Theoretical Prerequisites

When texts are not coherent, they do not make sense or they make it difficult for the reader to understand (Halliday and Hassan, 1992). In order to make the speech or written text coherent, consistent, easy to follow and understandable we can use *cohesive signposts in discourse* – discourse markers (Granger, 1996, p.80). Discourse markers have been called differently in literature. The terms *discourse particle* (Schourup, 1983; Fisher and Gruyter, 2000), *connective* (Salkie, 1995; Axelrod and Cooper, 2001; Celle and Huart, 2007), *insert/ discourse marker* (Biber et al., 1999), *connector* (Copage, 1999; Stephens, 1999; Frodesen and Eyring, 2000), *discourse marker/ utterance indicator/ filler* (Pridham, 2001), *linker* (Foley and Hall, 2003), *pragmatic marker/ discourse marker* (Aijmer, 2004; Aijmer and Simon-Vandenberg, 2006; Carter and McCarthy, 2006) were used. Such scholars as Fisher and Gruyter (2000, p.12) who investigated functional polysemy (multifunctionality) of discourse markers state that

the great number of different descriptive terms for this heterogeneous group indicate that firstly there is no single accepted word class definition, and that secondly the terms chosen depend very much on the perspective under which discourse particles are studied.

For example, Salkie (1995), Axelrod and Cooper (2001), and Celle and Huart (2007) use the term *connective* to mean an element used for linking (a word or phrase which indicates a connection between parts of a text), while Pridham (2001, p.30) uses the term *utterance indicator* to mean a word which features the structure of the conversation. It is noteworthy that the term *discourse marker* was chosen as the most general term in the present paper for items which help to structure the discourse. Discourse markers in the present paper include words belonging to different parts of speech such as conjunctions, particles, adverbs, interjections etc. Discourse markers are defined and classified differently by various scholars. The present paper relies on the definition and classification of discourse markers provided by Carter and McCarthy. Firstly, the scholars divide discourse markers into *single words and miscellaneous items* and *phrasal or clausal*

items. Secondly, they separate different functional categories of discourse markers. Further, Carter and McCarthy (2006) separate a group of words called *linking adjuncts* “which indicate the semantic relationship between two clauses or sentences or paragraphs”. Moreover, Carter and McCarthy (2006) indicate that among all the discourse markers (both discourse markers and linking adjuncts) there are discourse markers more common in informal spoken language. Sometimes they are used in written texts to imitate a spoken style. Among such discourse markers the scholars mention: single words and miscellaneous items such as *anyway, yeah, cos, fine, good, great, like, now, okay, right/ (all)right, so, and, well, then, hey, ah, oh, look, listen, remember, incidentally, meantime, anyhow, only* and phrasal or clausal items such as *you know, you see, I mean, as I say, for a start, mind you, just think, as I was saying, as it were, if you like, in a manner of speaking, in other words, in general, speaking of which, not to say, or rather, so to speak, strictly speaking, that's to say/ that is to say, to put it another way, to put it bluntly/ mildly, by the way, there you go, at the end of the day, talking about, while I think of it, as well, on top of it all, to cap it all, to crown it all, what's more/ what is more, then again, mind you, for a start.*

Biber et al. (1999, p.1085) state that “discourse markers are often ambiguous because they share the function of a discourse marker with an adverbial function”. Yet, Carter and McCarthy (2006, p.208) provide the difference between discourse markers (which appear outside of the clause) and other items used within the clause which should not be mixed: e.g. *I didn't really need it but I bought it anyway*. There is an in-clause use in this example; therefore, *anyway* does not serve as a discourse marker. Another example provided by Carter and McCarthy (ibid.) show *anyway* as outside of the clause structure (functioning as a discourse marker). For example:

A: *But you only pay one way.*

B: *Oh do you?*

A: *Yeach you only pay going into Wales. You don't pay coming out.*

B: *Oh. Right.*

A: *But er yeach, anyway, we drove in the rain and the dark for eight hours.*

Discussion of the Results

Table 1. Spoken Discourse Markers in *LICLE*, *LOCNESS-BR* and *LOCNESS-US* Corpora.

Corpora	Discourse markers	Per 10000 w
<i>LICLE</i>	358	23
<i>LOCNESS-BR</i>	108	9
<i>LOCNESS-US</i>	167	10

The investigation revealed that discourse markers more typical of speech than of academic writing were used by both the Lithuanian learners and the native learners. Out of the 63 types of spoken discourse markers 30 were found in *LICLE*, *LOCNESS-BR* and *LOCNESS-US* corpora. As Table 1 illustrates, the number of occurrences of the

spoken discourse markers in *LICLE*, *LOCNESS-BR* and *LOCNESS-US* corpora significantly differs. Since the three corpora differ in size, the number of tokens per 10 000 words is provided in order to give normalized frequency of discourse markers.

The frequency counts indicate that the Lithuanian learners used spoken discourse markers more frequently than the native learners. The comparative analysis showed that the spoken discourse markers found in *LICLE* are more similar to the spoken discourse markers found in *LOCNESS-US* than in *LOCNESS-BR*. It might be due to the fact that the American learners use more spoken discourse markers than the British learners. Even though spoken discourse markers were found in all the three corpora, the investigation suggests that the native speakers make a better distinction between the discourse markers used in academic essay writing and conversation. The most likely explanation is that the Lithuanian learners are less familiar with spoken discourse markers. It supports Granger's (1998, p.80) investigation which showed that learners have problems with differentiation between discourse markers used in conversation and academic essay writing. According to this scholar, "one problem for learners is that the use of connectors is sensitive to register and discourse type". One of the reasons could be that the learners' course books and other materials they use, lack of some important information about discourse markers. The requirements to use a formal style are indicated and among the items making the style more formal discourse markers are included; however, examples are indicated with no explanations on stylistic and statistical grounds of discourse markers. For example, *first*, *firstly* (to begin an essay) or *then*, *subsequently* (for middle steps) are given without indicating that *firstly* is more formal than *first*, and *then* is slightly informal, while *subsequently* is formal (Granger, 1998, p.175).

Discourse Markers that Mark the Same Functional Categories

Another interesting question is the difference in the number of spoken discourse markers used in *LICLE*, *LOCNESS-BR* and *LOCNESS-US* corpora to mark the same functional categories. Spoken discourse markers appeared to be of 15 functional categories (see Table 2).

The distribution of the functional categories of spoken discourse markers differs. The main similarity is that *sequencing* relations are most common in all the three corpora. It suggests that learners use spoken discourse markers mostly to indicate the order in which things occur (Carter and McCarthy, 2006, p.216). Resultative discourse markers take the second place in terms of frequency in *LICLE* and *LOCNESS-BR*; whereas, this place is taken by reformulation discourse markers in *LOCNESS-US*. Six of the functional categories were used more frequently in *LICLE* in comparison to *LOCNESS-BR* and *LOCNESS-US*. As the results show the Lithuanian learners used more of *inference*, *sequencing*, *resultative*, *reformulation*, *resuming* and *time* discourse markers than the native learners. *Response*, *difficulty to formulate*, and *diverting* relations were rarely used in *LOCNESS-US* and could not be found at all in *LICLE* and *LOCNESS-BR*. It was chosen to look at the frequency of individual spoken discourse markers of the

four functional categories more frequent in *LICLE* in comparison to *LOCNESS-BR* and *LOCNESS-US* in order to find out the major differences.

Table 2. Distribution of Functional Categories of Spoken Discourse Markers.

Functional Category	LICLE	Per 10 000w	LOCNESS-BR	Per 10 000w	LOCNESS-US	Per 10 000w
<i>Sequencing</i>	136	8,77	27	2,43	61	3,63
<i>Resultative</i>	87	5,61	31	2,79	24	1,43
<i>Reformulation</i>	35	2,90	15	1,35	25	1,60
<i>Concessive</i>	30	1,94	19	1,71	10	0,59
<i>Resuming</i>	25	1,61	5	0,45	20	1,19
<i>Inference</i>	23	1,48	3	0,27	1	0,06
<i>Opening</i>	17	0,45	6	0,54	16	0,83
<i>Focusing attention</i>	3	0,19	0	0,00	5	0,30
<i>Monitoring shared knowledge</i>	1	0,00	0	0,00	1	0,06
<i>Time</i>	1	0,06	0	0,00	0	0,00
<i>Additive</i>	0	0,00	2	0,18	0	0,00
<i>Response</i>	0	0,00	0	0,00	1	0,06
<i>Difficulty to formulate</i>	0	0,00	0	0,00	1	0,06
<i>Diverting</i>	0	0,00	0	0,00	1	0,06
<i>Closing</i>	0	0,00	0	0,00	1	0,06
TOTAL:15	358		108		167	

Inference. Inference discourse markers are much more common in the Lithuanian learners' academic essays in comparison to the native learners' essays. It confirms the investigation made by Biber (1999, p.886) which revealed that inference discourse markers are one of the most frequent items in conversation. The present investigation showed that *then* is the only inference discourse marker used in the three corpora (see Appendix 1). Again, it supports the investigation of linking adverbials across registers made by Biber et al. (1999, p.883) which revealed that high frequency of "spoken linking adverbials is due to the relatively high frequency of *then* as an inference linking adverbial". This discourse marker can perform various functions. As the results show, it was used as a resultative, time, summative, listing and inference discourse marker. When it is used as an inference discourse marker it is treated as an informal spoken discourse marker (Carter and McCarthy, 2006, p.260). While *then* serves as the fourth most common spoken discourse marker in *LICLE*, it is hardly ever used in the native learners' essays. It suggests that the Lithuanian learners are less familiar with *then* as a spoken discourse marker. At first glance this seems to be due to the influence of their mother tongue

(Lithuanian) where the inference discourse marker *then* (Lith. *tuomet*) is commonly used. One more explanation is that the information presented in different sources is contradictory. Some scholars state that *then* is usually used in more formal English, others include it in the list of the most frequent discourse markers in spoken English. These contradictory statements could be due to the different functions that this discourse marker performs. However, the scholars do not indicate full classification or explanation of these discourse markers. Such information can be misleading. For example:

1) *Postmodernists abandoned the notion of history completely relying on historiographic metafiction. They challenge history, stating that it is transferred through language, and, therefore, already is as interpretation. People do not know the real history, and any opinion can be a part of a real history then* (LICLE).

2) *The woman does not accept she alone was responsible for her inaction, and says there was nothing she could do. This, then, is Bad Faith, as it denies both choice and responsibility* (LOCNESS-BR).

As illustrated by the examples above, discourse marker *then* “marks one idea as an inferred result of another” (Biber et al., 1999, p.878). The most likely suggestion is to use *therefore* (which is a formal resultative discourse marker) instead of *then*. Moreover, the present investigation showed that *then* is commonly used in questions. As Granger (1998, p.105) argues, “an overuse of questions can reduce their argumentative value and increase the often more informal style of the writing”. For example:

3) *The majority of them are of working age seeking for comfort and profit, fleeing from unemployment, low wages and social insecurity. Some blame them of deception, unwillingness in helping the nation while it is still fragile in every aspect. Where are their patriotic feelings then?* (LICLE).

Sequencing. Sequencing spoken discourse markers take the second place in terms of more frequent functional categories used in LICLE in comparison to LOCNESS-BR and LOCNESS-US (see Table 2). The present investigation revealed that compared to the overall number of *and* in LICLE, LOCNESS-BR and LOCNESS-US (246 vs 190 and 155 accordingly) almost half appeared to be more of spoken nature in LICLE and only a small part in the native learners’ essays (see Appendix 1). This discourse marker is remarkable in that it appeared in LICLE to be seven times more frequent than in LOCNESS-BR and more than twice as common as in LOCNESS-US. It confirms Granger’s (1998, p.137) investigation on *sentence- initial trigrams* which showed that learners have higher tendency to start their sentences with initial *And*. For example:

4) *Therefore, if a future world were to use only one language who is to decide which language it should be? And is it at all possible to decide?* (LICLE).

5) *Can one really justify an innocent child being denied a chance at life as moral? And is it moral for doctors who have been sworn to abide by the Hippocratic Oath to turn*

their backs to that oath and perform abortions? (LOCNESS-US).

So is another sequencing discourse marker more common in speech than in writing (see Appendix 1). It is one of the most multifunctional discourse markers. As the present investigation revealed it can have resultative, resuming and sequencing relations. Sequencing *so* is not as common in spoken English as the resultative *so*, but it is still treated as more of spoken nature (Carter and McCarthy, 2006, p.216). The results indicate that sequencing *so* was nearly four times as frequent in LICLE in comparison to LOCNESS-BR and LOCNESS-US. It could be due to the teaching materials which “concentrate more on *global* rather than *local* aspects, without indicating that the repeated use of *so* is regarded as bad style” (Carter and McCarthy, 2006, p.176). For example:

6) *History becomes a topical issue especially during the postmodernist epoch. So, this brings me to the discussion of differences of the two traditions* (LOCNESS-BR).

In the example above sequencing *so* was used to mark how one thing leads to another, to start an idea. It creates an impression that the writer plans what to write next and makes a pause after *so*. This feature is very common in spoken English. It is called *a hesitation pause* by Biber et al. (1999, p.1053). It is a short period of silence during which the speaker thinks what to say.

Sequencing discourse marker *in general* was the most frequent to appear in the British learners’ essays and was least commonly used in the American learners’ essays (see Appendix 1). It suggests that the learners (particularly British and American) tend to use different spoken discourse markers. It could be due to the fact that different discourse markers are popular in British spoken English and American spoken English; whereas, the Lithuanian learners are taught to use British English which explains their similar tendency to use this discourse marker.

7) *Should scientists bear the moral responsibility for their work? In general scientist **don't actually** know what they have discovered* (LOCNESS-BR).

8) *In general it is possible to say that modernists and postmodernists have more similarities in literature than differences* (LICLE).

As illustrated in the example 7, discourse marker *in general* is preceded and followed by other spoken-like items: a question (which should be used economically), contraction *don't* (which is not appropriate in an academic essay writing), and emphatic item *actually* called *attitudinal disjunct* by Granger (1998, p.84), which is used in order to emphasize the truth of scientists’ ignorance. It seems that the student is trying to create an impression which may lead to overstatement.

Resultative is the third functional category of the spoken discourse markers which is more frequent in the Lithuanian learners’ essays in comparison to the native learners’ academic essays. *Resuming* relations are slightly more common in LICLE too. This higher frequency is due to a relatively high frequency of *so* in learners’ essays (see

Appendix 1). It shows that the Lithuanian learners tend to use *so* to express a result or to resume an interrupted or diverted topic more often than the native learners. For example:

9) *When an area of a city becomes known for a high crime rate the result is a drop in the real estate value for that area. So even though an individual can try to add value to his or her real estate possessions in these areas the effect of crime will cause a direct loss of money for the law abiding citizen (LOCNESS-US).*

10) *When a person is about to reconcile with his servitude, he kills the monster and flings him to the public. So, does writing really make people suffer or is it a joyous adventure taken with easiness? (LICLE).*

The example 9 illustrates the usage of a resultative *so*; whereas, the resuming *so* is illustrated in the latter example. A higher frequency of the resultative and resuming *so* in the Lithuanian learner academic essay writing in comparison to the native learners, could be due to the process of teaching where grammatical accuracy is treated as more important than stylistic accuracy. Most likely suggestion is to use *therefore* instead of the resultative *so*.

It is noteworthy that the discourse marker *of course* is commonly used to express a result in the learners' academic essays too. As the present investigation showed, sometimes it is used as a concessive discourse marker to indicate contrast. *Of course* is more common in the Lithuanian learners' essays when it functions both as a resultative and concessive discourse marker. For example:

11) *They wrote despising all nouns, and I love them. Of course, they probably are too professional to compare to, however, they make good examples to my statement (LICLE).*

12) *However, even students of that kind may miss the chance to study only because the financial problems. Of course, there is a possibility to take a loan but the system of getting loans has to be reformed as well in order to make it easier to get and pay back (LICLE).*

Of course is discussed in Granger (1998, p.128), Gilquin and Paquot (2007, p.4) as one of the most frequent speech-like adverbs emerging in learner academic writing. The investigation shows that the students lack of information about the emphatic items. If used too many, they make their writing spoken-like.

Reformulation. The fourth functional category of the spoken discourse markers which is more frequent in the Lithuanian learners' essays in comparison to the native learners' is *reformulation*. It is important to mention that some of the items of this functional category are more informal (e.g. *well, by the way, I mean*) and others are more formal but frequently found in conversations. The distribution of spoken reformulation discourse markers varies in each corpora. *In other words* and *well* are two most common spoken discourse markers of this functional category in LICLE, *that is to say* and *not to say*-in LOCNESS-BR, and *well, in other words*-in LOCNESS-US. The investigation suggests that while the Lithuanian learners

use *in other words*, the British learners tend to use *that is to say* in order to reformulate an expression. For example:

13) *Unfortunately, money cannot multiply by themselves and must be taken from somewhere. **So** are these problems worth taking them from others or, in other words, do we need a financial reform in our University? (LICLE).*

14) *Sartre writes this play with the basic theme of socialism running through all the scenes. That is to say he intends to illustrate their struggle to gain power. It is true to say that at no stage **does** Sartre **actually** criticise this socialism openly and that owing to the internal structure one may be led to believe that the play is openly supporting the cause; however, **this** I do not feel is the case (LOCNESS-BR).*

As the examples above illustrate, other spoken features tend to occur while restating an expression in the learners' essays. Interestingly, the discourse marker *so* was used in the Lithuanian learner's essay (see example 13). The example 14 includes lexical items *peculiar to spoken language* (Aijmer, 2004, p.173). Emphatic *does* and *actually* were used. Moreover, a deictic item *this* which is usually used to refer to things which are close in space and time (Carter and McCarthy, 2006, p.178) occurred in the same essay too. It is noteworthy that *this* is fronted in the sentence which means that it is highlighted. "Fronting may be used to emphasize what the speaker considers to be especially significant" (Carter and McCarthy, 2006, p.192).

The second most common informal reformulation discourse marker in LICLE is *well*. As indicated in Appendix 1, it is also the most common discourse marker of this functional category in LOCNESS-US. *Well* is mostly used in conversation to signal a shift in the direction of the discourse, except rare occurrences in informal writing (Carter and McCarthy, 2006; Biber et al., 1999). For example:

15) *Allowing same-sex marriages is not such a bad idea as most of Lithuanians think. This would even solve some of the country's problems. The only question remains: is the country ready for that? Well, **I think** not quite yet (LICLE).*

16) *Well let's be tough on law and order by cracking down on criminals, but not doing it by committing another crime, and murdering someone because they made a huge mistake (LOCNESS-US).*

In the example 15 *well* is followed by *I think* "which is pervasive in informal conversation" (Aijmer, 2004, p.176). It signals that the speaker is uncertain "indicating consultation by the speaker of his or her current thoughts" (Schourup, 1983, p.64). The present investigation showed that *I think* is commonly used in all the three corpora. It supports the investigation made by Aijmer (2004, p.184) with Swedish learners and native speakers which showed that *I think* was used by both the Swedish learners and the native learners (only in different positions of the sentence). Moreover, *well* creates a space for planning what to say. It is preceded by a question which creates an impression of a dialogue. The example 16 illustrates *well* used before the first person imperative *let's*. Discourse marker *well* serves as an *utterance launcher* (Biber et al., 1999, p.1118). As

stated by this scholar, *well* is the most common utterance launcher in American English.

The third most common reformulation discourse marker in *LICLE* is *I mean*. For example:

17) *Moreover, it seems to be impossible as the language of each nation is very much culture-dependent. I mean, each culture has its own perception on the world which can hardly be changed (LICLE).*

18) *I mean if the model in the commercial can look like that because she uses that certain product — so can I (yeah right) (LOCNESS-US).*

As illustrated by the examples, “*I mean* redirects the ongoing talk by introducing *modifications* which both correct and add to the previous contribution” (Schourup, 1983, p.106). In the example 17 the learner uses *I mean* in order to clarify and explains what *culture dependent* means. In the example 18, the writer indicates such lexical items as *yeah* and *right*. *Yeah* is used to focus listener’s attention and *right* in order to facilitate closing or pre-closing (Carter and McCarthy, 2006, p.215). These examples create an image of unplanned speech where repetitions and reformulations are common.

Such lexical items as *or rather*, *by the way*, *so to speak*, *as it were* and *not to say* were rarely found in the three corpora. For example:

19) *However, such a rebellion cannot be seen clearly in each minority work, and, therefore, the products of ethnic American literature cannot be categorized as merely the result of years of oppression. Or rather, this ever-changing and ever-challenging aspect of minority literature creates an especially important necessity that each work be considered individually as both a product of years of struggle and a work inherently distinct from any other (LOCNESS-US).*

20) *He is bleeding and quite shaken up, but he will definitely live. Oh, by the way the robber that was told to “run like **hell**” eventually got caught around the corner so when the mad gunman ran even faster to get away after he shot the little boy he too was apprehended (LOCNESS-US).*

21) *The research in the last decade or so has made unbelievable progress at every turn, not all of it ‘morally viable’, so to speak (LOCNESS-BR).*

22) *Getting rid of this part is, as it were, putting oneself under conditions of servitude, restricting the human natural ways of knowing (LICLE).*

23) ***In other words**, through the late realization of his guilt versus his peoples innocence, a certain sympathy is evoked for his misunderstanding of the way in which one should live alongside the concept of the absurd. Not to say, that we see his actions as justified but that we do perceive a misinterpretation of the truth in life (LOCNESS-BR).*

The present investigation revealed that sometimes more than one spoken discourse marker appears in one paragraph of the essay. It might suggest that certain spoken discourse markers serve as an individual feature (frequently used by the same person). *By the way* is preceded by the

diverting discourse marker *oh* in the example 20. This item strengthens the oral tone of the essay as it is generally used to signal *an unexpected diversion in the conversation* (Carter and McCarthy, 2006, p.219). As the investigation made by Biber et al. (1999, p.1083) showed, it is the most commonly used *interjection* in a conversation with a function *to convey some degree of surprise, unexpectedness, or emotive arousal*. Other examples also show that learners tend to use not one spoken-like item in a paragraph of the essay. Such items as deictic *we* (see example 23) or a swearword *hell* (see example 20) occurred in the native learners’ essays. It is suggested to avoid using such items at all.

Conclusions

The findings showed that discourse markers more typical of speech than of academic writing occurred in both the Lithuanian learners and the native learners academic essays. Yet, the contrastive analysis of the spoken discourse markers in the three corpora revealed that the spoken discourse markers were more frequently used by the Lithuanian learners in comparison to the native learners. The higher frequency of the spoken discourse markers in *LICLE* might be determined by a number of factors: students’ course books which lack stylistic suggestions of discourse markers, the contradictory information provided in different sources, the communicative approach to the second language teaching which is based on spoken English (interactive activities such as dialogues are frequently used in the classroom) that enables students to focus on spoken language which influences their academic (essay) writing.

Furthermore, this study revealed that out of the fifteen functional categories of the spoken discourse markers found in the three corpora the Lithuanian learners used much more of inference (*then*), sequencing (*and, so*), resultative (*so, of course*) and reformulation (*in other words, talking about*) discourse markers in comparison to the native learners.

During the analysis of the spoken discourse markers other spoken features were used by both the Lithuanian learners and the native learners too. This might suggest that register confusion could be due not only to the language learning (English as a second language) but also to the learning how to write.

References

1. Aijmer, K., 2004. Pragmatic Markers in Spoken Interlanguage. *Nordic Journal of English Studies* 3(1). Gothenburg: Gothenburg University Publications, pp.173–190.
2. Aijmer, K., Simon-Vandenberg, A. M., 2006. *Pragmatic Markers in Contrast*. Amsterdam: Elsevier.
3. Aijmer, K., Altenberg, B., Johansson, M. (ed.), 1996. *Languages in Contrast*. Lund: Lund University Press, pp.37–51.
4. Axelrod, R. B., Cooper C. R., 2001. *The St. Martin’s Guide To Writing*. New York: St. Martin’s Press.
5. Baker, P., 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
6. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. *Longman Grammar of Spoken and Written English*. Harlow and London: Pearson Education Limited.

7. Carter, R., McCarthy, M., 2006. Cambridge Grammar of English. Cambridge: Cambridge University Press.
8. Celle, A., Huart, R., 2007. Connectives as Discourse Landmarks. Amsterdam: John Benjamins Publishing Company.
9. Copage, J., 1999. First Certificate Writing. Harlow: Pearson Education Limited.
10. Fisher, K., de Gruyter, M., 2000. From Cognitive Semantics to Lexical Pragmatics. Berlin, New York: Mouton de Gruyter.
11. Foley, M., Hall, D., 2003. Advanced Learner's Grammar. Harlow: Pearson Education Limited.
12. Frodesen, J., Eyring, J., 2000. Grammar Dimensions. Boston: Heinle & Heinle Publishers Inc.
13. Gilquin, G., Paquot, M., 2007. Spoken Features in Learner Academic Writing: Identification, Explanation and Solution. Proceedings of the 4th Corpus Linguistics Conference, University of Birmingham, 27–30 July 2007. Available at: http://ucel.lancs.ac.uk/publications/CL2007/paper/204_Paper.pdf [assessed 22 November 2010]
14. Granger, S., 1996. From CA to CIA and Back: an Intergrated Approach to Computerized Bilingual and Learner Corpora. Sweden: Lund University Press.
15. Granger, S., 1998. Learner English on Computer. London: Longman.
16. Granger, S., Hung, J., Tyson, S. P., 2002. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Amsterdam: John Benjamins Publishing Company.
17. Halliday, M. A. K., Hasan, R., 1992. Cohesion in English. London: Longman group Limited.
18. Halliday, M. A. K., Teubert, W., Yallop, C., Čermakova, A., 2004. Lexicology and Corpus Linguistics. New York, London: Continuum.
19. Huddleston, R., Pullum, G. K., 2002. The Cambridge Grammar of the English Language. Cambridge: Cambridge University Press.
20. Kennedy, G., 1998. An Introduction to Corpus Linguistics. London: Longman group Limited.
21. Leech, G., Svartvik, J., 1994. A Communicative Grammar of English. London: Longman Group Limited.
22. Masaitienė, D., 2003. Discourse Markers in English and Lithuanian. Kalbotyra, nr. 53 (3). Vilnius: Vilniaus universiteto leidykla, pp.64–70.
23. Meyer, C. F., 2002. English Corpus Linguistics. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511606311>
24. Pridham, F., 2001. The Language of Conversation. London: Routledge.
25. Salkie, R., 1995. Text and Discourse Analysis. London: Routledge Taylor & Francis Group.
26. Schourup, L. C., 1983. Common Discourse Particles in English Conversation. Columbus: The Ohio State University.
27. Stephens, M., 1999. Proficiency Writing. Essex: Pearson Education Limited.

Alė Šimčikaitė

Šnekamosios kalbos diskurso žymikliai anglų kalbą besimokančiųjų akademinuose rašiniuose

Santrauka

Šiame darbe analizuojami šnekamosios kalbos bruožai anglų kalbą besimokančiųjų akademinuose rašiniuose pasirenkant diskurso žymiklių analizę. Diskurso žymikliai plačiai vartojami rašytinėje ir sakytinėje kalboje teksto rišlumui sustiprinti, tačiau mažai dėmesio skiriama individualiems diskurso žymikliams, kurie gali būti formalūs, pusiau formalūs ir neformalūs, bei atitinkamai vartojami formalioje (rašytinėje) arba neformalioje (šnekamojoje) kalboje. Neformalūs diskurso žymikliai panaudoti akademinuose rašiniuose nulemia stilistiškai netinkamą rašinio toną. Šiuo tyrimu siekiama išsiaiškinti, kokių mastu diskurso žymikliai, kurie traktuojami kaip labiau paplitę šnekamojoje kalboje, yra naudojami pažengusiųjų lietuvių bei gimtakalbių (britų ir amerikiečių) studentų akademinuose rašiniuose. Taip pat atkreipiamas dėmesys į kitus šnekamosios kalbos bruožus, pastebėtus studentų rašiniuose analizuojant diskurso žymiklius. Tyrimo duomenys gauti iš trijų tekstynų: LICLE, kuris yra sudarytas iš lietuvių studentų rašinių (parašytų anglų kalba) ir susideda iš 154 992 žodžių, LOCNESS-BR – gimtakalbių britų (susideda iš 111 127 žodžių) ir LOCNESS-US – gimtakalbių amerikiečių (sudarytas iš 168 231 žodžių). Tyrimui atlikti buvo taikomi kiekybinis, kokybinis ir lyginamosios analizės metodai. Tyrimo rezultatai parodė, kad lietuvių mokiniai yra linkę naudoti daugiau šnekamosios kalbos diskurso žymiklių nei gimtakalbiai mokiniai. Tokių rezultatų priežastys gali būti įvairios: per mažas dėmesys skiriamas stilistiniams diskurso žymiklių bruožams (pvz., formalūs arba neformalusis diskurso žymiklis), prieštaringa informacija pateikiama įvairiuose šaltiniuose apie diskurso žymiklių vartojimą, plačiai taikomas komunikacinis anglų kalbos mokymo metodas (pagrįstas bendravimu, kalbėjimu, diskusija), kuris sutelkia mokinių dėmesį į šnekamąją kalbą, kuri ir atsispindi jų rašiniuose. Gimtosios lietuvių kalbos įtaka nežymiai pastebima analizuojant šnekamosios kalbos diskurso žymiklius. Penkiolika funkcinių kategorijų buvo išskirta analizuojant šnekamosios kalbos diskurso žymiklius, jos padėjo atskleisti, kurias funkcijas atliekantys diskurso žymikliai yra naudojami dažniau lietuvių mokinių rašto darbuose nei gimtakalbių rašiniuose. Taip pat buvo pastebėta nemažai kitų sakytinės kalbos bruožų, kurie suteikia šnekamosios kalbos toną akademiniam rašiniams.

Straipsnis įteiktas 2012 01
Parengtas spaudai 2012 05

About the author

Alė Šimčikaitė, teacher, The General Jonas Žemaitis Military Academy of Lithuania, Institute of Foreign languages; asist., Vilnius University, Institute of Foreign languages.

Research interests: discourse analysis.

Address: Military Academy of Lithuania, Šilo str. 5 a., LT-10322 Vilnius, Lithuania/ Vilnius University, Universiteto str. 3, LT-01122 Vilnius, Lithuania.

E-mail: ale.simcikaitė@yahoo.com

Appendix

Appendix 1. Distribution of the Spoken Discourse Markers per Functional Category.

FUNCTIONAL CATEGORY	SPOKEN DISCOURSE MARKER	LICLE	LOCNESS-BR	LOCNESS-US
<i>Sequencing</i>	<i>And (sent. initial)</i>	109	15	51
	<i>So</i>	23	6	8
	<i>In general</i>	3	5	1
	<i>Going back to</i>	1	1	1
TOTAL:	4	136	27	61
<i>Resultative</i>	<i>So</i>	59	21	21
	<i>Of course</i>	28	10	3
TOTAL:	2	87	31	24
<i>Reformulation</i>	<i>In other words</i>	19	1	4
	<i>Well</i>	9	1	15
	<i>I mean</i>	3	0	3
	<i>Or rather</i>	1	1	1
	<i>By the way</i>	1	0	1
	<i>So to speak</i>	1	1	0
	<i>As it were</i>	1	0	0
	<i>That is to say</i>	0	7	0
	<i>Not to say</i>	0	3	1
	<i>Incidentally</i>	0	1	0
TOTAL:	10	35	15	25
<i>Concessive</i>	<i>Of course</i>	30	19	10
TOTAL:	1	30	19	10
<i>Resuming</i>	<i>So</i>	25	5	20
TOTAL:	1	25	5	20
<i>Inference</i>	<i>Then</i>	23	3	1
TOTAL:	1	23	3	1
<i>Opening</i>	<i>Now</i>	7	6	14
	<i>Talking about</i>	10	0	2
TOTAL:	2	17	6	16
<i>Focusing attention</i>	<i>Remember</i>	3	0	1
	<i>Look</i>	0	0	2
	<i>Just think</i>	0	0	1
	<i>Yeah</i>	0	0	1
TOTAL:	4	3	0	5
<i>Monitoring shared knowledge</i>	<i>You know</i>	1	0	1
TOTAL:	1	1	0	1
<i>Time</i>	<i>Meantime</i>	1	0	0
TOTAL:	1	1	0	0
<i>Additive</i>	<i>What's more</i>	0	1	0
	<i>To cap it all</i>	0	1	0
TOTAL:	2	0	2	0
<i>Response</i>	<i>Fine</i>	0	0	1
TOTAL:	1	0	0	1
<i>Difficulty to formulate</i>	<i>Like</i>	0	0	1
TOTAL:	1	0	0	1
<i>Diverting</i>	<i>Oh</i>	0	0	1
TOTAL:	1	0	0	1
<i>Closing</i>	<i>Right</i>	0	0	1
TOTAL:	1	0	0	1
TOTAL:		358	108	167