REDOVNA SEKCIJA – REGULAR SECTION – ALLGEMEINER TEIL

ČLANCI – ARTICLES – ARTIKEL

Jolanta Šinkūnienė
Vilnius University

Reformulation markers in academic discourse: A cross-linguistic and cross-disciplinary study[*]

Using quantitative and qualitative approaches alongside contrastive analysis, this paper investigates distribution frequency and functions of reformulation markers in academic discourse in two languages (English and Lithuanian) and three science fields (humanities, medicine, technology). The English language data is taken from the academic language sub-corpus of the Corpus of Contemporary American English, while the Lithuanian language data comes from the Corpus Academicum Lithuanicum, a specialised synchronic corpus of written academic Lithuanian. The results show that it is the humanities scholars who employ reformulation markers most frequently in both languages. They also employ a wider range of reformulation markers and use them in more diverse ways than scholars in the hard fields. The most frequent function of reformulation markers irrespective of language and science field is the interpretation of explicit content. The analysis highlights the importance of the discipline and genre in the distribution and use of reformulation markers.

Key words: reformulation markers; academic discourse; corpus-based; English; Lithuanian.

1. Introduction

Academic discourse in different science fields and cultures has become one of the favored objects of applied linguists in the past several decades. One area of academic discourse that has attracted considerable attention of scholars is the use of metadiscourse markers, which are employed by academic writers to create convincing, powerful, and coherent discourse. A growing body of literature on linking words, hedges, boosters, attitudinal lexis, personal pronouns, and other metadiscourse elements has revealed interesting ways in which the "big culture" (i.e. national culture) and the "small culture" (i.e. disciplinary culture) (cf. Atkinson 2004) play a role in the creation of academic text (see Hyland 2005; Fløttum et al. 2006; Mur-Dueñas 2011; Hyland & Sancho Guinda 2012, *inter alia*). These studies have also contributed to a better understanding of what metadiscourse is and how its expression and features vary from culture to culture and from discipline to discipline.

Metadiscourse has generally been understood as "discourse about discourse or communication about communication" (Vande Kopple 1985: 83). Traditionally following the distinction of macro-functions of language proposed by Halliday (1973), metadiscourse markers have been classified into textual and interpersonal (Vande Kopple 1985; Crismore et al. 1993, *inter alia*). The primary function of textual markers is to indicate how "we link and relate individual propositions so that they form a cohesive and coherent text and how individual elements of those propositions make sense in conjunction with the other elements of the text in a particular situation" (Vande Kopple 1985: 87). Interpersonal markers help reveal "our personalities and our reactions to the propositional content of our texts and characterize the interaction we would like to have with our readers about that content" (Vande Kopple 1985: 87).

In his innovative model of metadiscourse proposed in 2005, Hyland rejects the classification of metadiscourse elements into textual and interpersonal suggesting instead that "all metadiscourse is interpersonal" (2005: 41). Hyland offers a two-dimensional model which classifies metadiscourse markers into two categories: interactive and interactional. The interactive dimension is concerned with the ways of organising discourse with the reader's needs in mind and, therefore, includes markers that show links between the ideas, offer explanations of propositional content, refer to different parts of the text or make reference to other literature. The interactional dimension provides the dialogic perspective to discourse, allowing authors to show their stance and commitment to propositions as well as involve the reader into the construction of the argument.

Interest in metadiscourse and its effect on text has opened the doors to a number of cross-disciplinary and cross-linguistic studies of academic discourse. It is especially the interactional dimension that has been thoroughly explored by various researchers: studies of the use of hedges, boosters, self-mention, attitude, and engagement markers revealed interesting insights into how scientific knowledge is presented in different disciplines and cultures (see, for example, Hyland 2001; Vassileva 2001; Mur-Dueñas 2007; Bondi 2008; Sanderson 2008). Empirical results show that "writers in different disciplines represent themselves, their work and their readers in different ways, with those in the humanities and social sciences taking far more explicitly involved and personal positions than those in the science and engineering fields" (Hyland 2008: 12). Cross-linguistic research, which has typically focused on metadiscourse use in English academic texts in comparison to texts in other languages, has suggested important differences in how scientific arguments are negotiated across cultures. Research on personal pronoun use, for example, suggests that in many cultures other than English, researchers avoid using *I*, choosing *we* and impersonal or passive structures instead (for an overview see Mur-Dueñas & Šinkūnienė 2016). Both cross-linguistic and cross-disciplinary insights derived from studies on interactional metadiscourse contribute significantly to academic identity research as well as provide important EAP/ESP guidelines for novice or/and non-native English speaking researchers.

The interactive dimension markers, on the other hand, have received less attention in literature despite the fact that they play a significant role in successful communication of ideas in academic texts. One category of interactive metadiscourse markers, code glosses, are important contributors to "coherent, reader-friendly prose" (Hyland 2007: 266) as they are employed to facilitate the reader's understanding of the message the writer tries to convey. Code glosses signal explanations, illustration or elaboration of the propositional material and thus play an important rhetorical role in reflecting the author's view of the reader with regard to the amount of information s/he requires in order to understand the text. Hyland (2007) classifies code glosses into reformulation and exemplification markers according to the two main functions they perform in scientific texts. It is namely the first category of code glosses, i.e. reformulation markers, that are the object of the present study.

2. Theoretical background

Initially, reformulation has attracted attention in linguistic literature as an element of cohesion and apposition (Halliday & Hassan 1976; Biber et al. 1999). Introduc-

ing conjunctions as a cohesive device, Halliday and Hassan (1976: 227) emphasise the "function they have of relating to each other linguistic elements that occur in succession but are not related by other, structural means". Among the markers of the conjunctive relations of additive type Halliday and Hassan (1976: 248) indicate expository items *that is*, *that is to say*, *in other words*, *to put it another way* as frequently occurring in apposition function. In their classification of semantic categories of linking adverbials, Biber et al. (1999: 876) establish adverbials of apposition, which "show that the second unit of text is to be treated either as equivalent to or included in the preceding unit." Biber et al. note that such adverbials are more frequent in academic prose than in conversation (1999: 884).

Later studies of individual linguistic items, especially of such multifunctional elements as discourse markers or pragmatics markers revealed that the function of reformulation was frequently present within their diverse functional semantic repertoire. For example, research on the Swedish discourse marker *alltså* based on its translation correspondences (Aijmer 2007) shows that especially the paraphrastic *alltså*, among its other uses, frequently performs the function of clarifying or elaborating propositional meaning without changing the semantic content. However, it is also used for another type of reformulation "with translations such as *finally*, *in fact*, *anyway*, which are closer to the inferential marker but signal that the speaker distances him/herself from what is said." (Aijmer 2007: 54). In a similar manner it is noted that the Norwegian particle *altså* could be used to offer metapragmatic guidance to the hearer, especially in such pragmatically important situations when a conclusion has to be drawn (Vaskó & Fretheim 1997: 245). Such studies on multifunctional linguistic items suggest that reformulation is a complex linguistic phenomenon and the range of the markers that can actually perform the function of reformulation is by no means delimited, especially in spontaneous speech or fiction.

Some scholars have applied Relevance Theory (Sperber & Wilson 1986) in their research on the functional and semantic pragmatic potential of reformulation markers. Following Sperber and Wilson (1986), Blakemore (1993) argues that reformulation is a stylistic device the purpose of which is to create particular contextual effects. The decision to reformulate, according to Blakemore (1993: 119), is "a decision about style, which like all decisions about style, is constrained by the search for relevance". Murillo (2012) combines typologies proposed by various scholars, basing her classification on the process of utterance interpretation as proposed in Relevance Theory. According to it, reformulation markers are considered as procedural items, which guide "the inferences that the hearer/reader makes when s/he interprets utterances" (Murillo 2012: 76). As such, they perform a number of func-

tions that are not restricted merely to an exact paraphrase, but can introduce definitions, specification, background information, conclusions, etc.

Reformulation has also been investigated in specialised discourse. Here the functions of reformulation markers are usually linked specifically to the rhetorical goals of the writers, their expectations and evaluation of the audience's knowledge and expertise in the field. Hyland's (2007) study of reformulation and exemplification markers (referred to as *code glosses*) in hard and soft fields shows clear disciplinary differences in the preference of one or another group of markers. Scientists in the hard fields rely on reformulation more frequently, whereas humanities and social sciences scholars prefer exemplification. They also use a greater variation of the reformulation devices in their articles. Even though the functional range of reformulation markers is reflective of specific disciplinary practices, their use also indicates "the routine significance of elaborative code glosses in the argumentation practices of all disciplines" (Hyland 2007: 283). The investigation of reformulation markers in judicial discourse (Mazzi 2011) likewise reveals rich potential of these linguistic elements to contribute to the effectiveness of the rhetorical argumentation of judges.

Some studies of reformulation markers in academic discourse have a cross-linguistic dimension. Cuenca (2003) looks at reformulation markers in a corpus of linguistic academic texts in English, Catalan, and Spanish. The results of her research show that authors in English use fewer reformulation markers than Spanish writers, whereas texts in Catalan occupy a middle position with regard to the frequency of reformulation cases. English also makes more use of simple fixed forms, whereas Catalan and Spanish academic writers tend to employ complex markers of reformulation. Cuenca explains these results by referring to cross-linguistic differences governing the writing of expository texts. Spanish academic writing is content-oriented and therefore wordier than the English linear style of writing, so more digressions are present in the articles of Spanish scholars.

Murillo (2012), on the other hand, finds that in Business Management research articles writers in English use far more reformulation markers than writers in Spanish, with Spanish English texts showing no statistically significant differences in comparison to reformulation use in English L1 articles. The difference in the use of reformulation markers between L1 English and L1 Spanish is explained by the potentially different needs of the audience: L1 English articles are written for a wider international audience, whereas L1 Spanish articles are intended for a smaller national audience, which potentially could have more insider knowledge about the matters discussed in the papers, so less explanation and reformulation is necessary.

The functional properties of reformulation markers in L1 English and L1 Spanish are also different. English authors use reformulation markers for the interpretation of explicit meaning far more frequently than Spanish authors in their native language. It is interesting that Spanish scholars writing in English seem to adapt to some extent the English native speaker writing patterns.

Since cross-linguistic studies show interesting trends in the distribution and use of reformulation markers in academic discourse, more studies in languages other than English are necessary to see which patterns can be observed in different disciplines. The aim of the present paper is to re-evaluate the role of discipline and language in the frequency and functions of reformulation marker use in academic discourse. The study explores reformulation marker employment in written academic texts in two languages (English and Lithuanian) and three science fields (humanities, medicine, technology).

3. Data and methods

The study uses data from two big synchronic corpora, the Corpus of Contemporary American English (COCA) and the Corpus Academicum Lithuanicum (CorALit). The size of the COCA is more than 560 million words from spoken language, fiction, popular magazines, newspapers, and academic texts balanced and collected over the time span of 1990–2017. As the focus of this research is on academic language, only academic sub-corpora of the humanities, medicine and science/technology were used for the study. These sub-corpora of the COCA consist of texts from academic journals. CorALit is a specialised synchronic corpus of written academic Lithuanian containing roughly 9 million words in five broad science fields (the humanities, the social sciences, the biomedical sciences, the technological sciences, and the physical sciences). Several genres are represented in each sub-corpus of CorALit including research articles, monographs, textbooks, reviews (for a more detailed description of the main features of the CorALit compilation and design see Usonienė et al. 2011).

The humanities, the (bio)medical sciences, and the technological sciences represent a disciplinary range from soft to hard science fields and have therefore been selected for the present study. Table 1 shows the number of words in each sub-corpus.[1]

---

[1] The paper does not use a full dataset of CorALit, but only the part of the data which is made available online at http://www.coralit.lt/node/5

Table 1. The size of the sub-corpora used for the analysis

Science field	CorALit (LT)	COCA (EN)
(Bio)medicine	519 070	6.7 million
Humanities	417 031	11.9 million
Technology	275 654	14.1 million
Total:	1 211 755	32.7 million

Quantitative and qualitative approaches were employed alongside contrastive analysis to reveal the ways in which the two languages and the three science fields are similar or different with regard to the use of reformulation markers. Since all sub-corpora are of very different sizes, raw numbers have been normalised to 10 000 words to enable comparison of quantitative distribution of the markers under study in different science fields. In order to evaluate whether the frequency data are statistically significant, the log-likelihood test (LL) was occasionally used with the critical value of 3.84 or higher at the level of $p < 0.05$.

Following Murillo (2012), the study focused on the following reformulation markers in English: *i.e., that is*, *that is to say*, *in other words*, *namely*. The repertoire of the main reformulation markers in Lithuanian is very similar and included such markers as *tai yra / t.y.* 'that is / i.e.', *kitaip tariant / sakant / pasakius / kitais žodžiais tariant* 'in other words', *būtent* 'namely'.

The markers under study were automatically extracted from each sub-corpus and every instance of use was then checked manually to avoid irrelevant word combinations or cases of use that do not perform reformulation functions (cf. examples 1–3):

(1) *Rutininiam ligonių būklės sekimui plačiausiai taikoma pilvo organų echoskopija, nes **tai yra** pigus ir plačiai prieinamas tyrimas.* (B-CorALit)
'Abdominal ultrasound is widely used for the routine follow up on the condition of the patients since that is an inexpensive and very accessible analysis'.

(2) *Why **that is**, however, remains unclear.* (T-COCA)

(3) *Address all correspondence and requests for reprints to: Lars **I.E.** Oddsson, DrMedSci, NeuroMuscular Research Center, Boston University*. (M-COCA)

The quoted examples are coded with the first letter of the science field and the title of the corpus, thus texts from the humanities field are coded as H-CorALit / H-

COCA, from the (bio)medical sciences as B-CorALit / M-COCA and from the technological field respectively as T-CorALit / T-COCA. The two languages are referred to in the article as LT for Lithuanian and EN for English.

Following Murillo (2012), the functions of reformulation markers were divided into three broad categories: 1) interpretation of explicit content; 2) explication of conceptual knowledge; 3) interpretation of implicit content.

Under the broad category of interpretation of explicit content, Murillo (2012) includes identification of referents, specification of cataphoric elements, and explanation of propositional material. This category is exemplified by (4–6):

(4) *This way the specific conditions, **i.e.**, the wall, floor, lighting, etc., will be correctly incorporated into the photographic component.* (H-COCA)

(5) *Ilgą laiką eksploatuotose šiluminių elektrinių vamzdynų ir katilų pitinginės korozijos paliestose plieno srityse aptikti koroziją sukeliantys elementai, **t. y.** K, Ca, Si, S ir Cl.* (T- CorALit)
    'The steel parts of thermal electric pipes and boilers, affected by pitting corrosion, contained   elements that cause corrosion, i.e. K, Ca, Si, S and Cl.'

(6) *Balansuodami ant intelektualinio chuliganizmo ribos drįstame teigti, jog istoriografijos tyrinėtojai pas mus yra savotiškos pasakų pabaisos, **kitaip tariant** –lietuviškoje istorijos mokslo tradicijoje istorikų akademinė autorefleksija vis dar yra retenybė.* (H-CorALit)
    'Balancing on the border of intellectual hooliganism, we take the liberty to argue that researchers in historiography here are some kind of monsters from fairy-tales, in other words, in Lithuanian historiographic tradition academic self-reflection still remains rare.'

Example (4) illustrates an identification process, in which the reformulation marker *i.e.* introduces the referents of the "specific conditions". *T.y.* 'i.e.' in (5) specifies which exact chemical elements were found in the corrosive parts of the metal equipment. In (6) the author of the text uses a creative metaphoric reference to present researchers of historiography as monsters from fairy-tales. Naturally, such a comparison calls for an explanation, which is introduced by the reformulation marker *kitaip tariant* 'in other words'.

The second broad classificational category, explication of conceptual knowledge, is comprised of definition and denomination (Murillo 2012: 78–79), and can be exemplified by (7–8):

(7) *Kai kurie autoriai siūlo, kad ankstyva kasos karcinoma būtų laikoma carcinoma in situ, **t. y**. vėžys, neišpliteṣ iš kasos latako epitelio, arba karcinoma, mažesnė negu 1 centimetras*. (B- CorALit)
'Some authors suggest that carcinoma in situ, i.e. cancer that has not spread outside of pancreatic duct cells, or cancer less than 1 cm in diameter, should be viewed as early pancreatic cancer'.

(8) *Arguments for the substantive dismissal of this Duality link can be found in the case where the company tracks receivables only by amount (**i.e.**, a balance-forward system).* (T-COCA)

T.y. 'i.e.' in (7) is used to signal a definition of the specific medical term *carcinoma in situ*, which the author of the text considers necessary to define for the reader. In (8) by means of denomination the authors introduce a field specific term.

Finally, the third functional category is interpretation of implicit content. It consists of two types of functions performed by reformulation markers. The first one is when reformulation markers introduce a mathematical operation or quantification:

(9) *BVP padidėjo nuo 45,7 iki 82,0 mlrd. Lt, **t. y**. beveik 1,8 karto.* (T-CorALit)
'GDP increased from 45.7 to 82.0 billion Lt, **i.e.** nearly 1.8 times'.

In example (9) the author provides GDP increase numbers and with the help of *t.y.* 'i.e.' immediately introduces mathematical calculations to show how many times GDP actually increased to save the readers mental effort to calculate the numbers themselves.

The second type of function performed by reformulation markers in this category is to introduce a conclusion:

(10) *Before a time when medical researchers were expected to conduct randomized controlled trials, Moe had proven to his own satisfaction that bracing worked, but only if the condition was detected early before the curve became too severe (in his judgment, greater than 60 necessitated surgery) and before the adolescent patient had reached skeletal maturity. 43 For bracing to work, **in other words**, time was of the essence.* (M-COCA)

Example (10) provides a concluding, generalising remark, which summarises the information provided in the previous sentences.

4. Results and discussion

The first part of this section looks at the distribution and frequency patterns of the reformulation markers listed above taking into account disciplinary and language specific variation in their use. The second part follows the functional classification framework suggested by Murillo (2012; 2018) in order to reveal the primary functions these markers perform in different disciplines and research cultures.

4.1. *Frequency of reformulation markers in the two corpora*

Table 2 provides quantitative data on frequency distribution of reformulation markers in the three science fields and the two languages.

Table 2. Frequency of reformulation markers in the analysed sub-corpora

Sub-corpus	Raw frequency	f/10.000
(Bio)medical LT	232	4.5
Medical EN	2462	3.7
Humanities LT	393	9.4
Humanities EN	4900	4.1
Technology LT	242	8.8
Technology EN	4051	2.9
Total LT:	867	7.2
Total EN:	11 413	3.5

The results in Table 2 show that Lithuanian scholars use more reformulation markers than scholars writing in English, i.e. the normalised frequency per 10 000 words is 7.2 vs. 3.5 respectively. As confirmed by the log likelihood test score, the difference in the use of reformulation markers in all three science fields is statistically significant with regard to the language in which texts were written. These results are similar to the findings in Cuenca's (2003) study of reformulation marker use in linguistic research articles in Spanish, Catalan, and English, which showed that Spanish authors used far more reformulation markers than the English authors. Cuenca suggests that this difference could be due to the different cultural styles of writing: the form-oriented, more linear, and allowing less digressions English style vs. the Spanish style, which is more content-oriented and not necessarily linear in text development. Previous comparative studies of Lithuanian and English academic discourse focusing on hedging (Šinkūnienė 2011) and personal pronoun use (Šinkūnienė 2018) demonstrated that these metadiscourse elements were employed differently in the two languages pointing to diverse rhetorical choices scholars writing in Lithuanian and in English make. Though it is tempting at this point to sug-

gest that Lithuanian and English rhetorical writing traditions may play a role in the frequency patterns of reformulation marker use as well, we will see later that other factors may also be important.

The distribution of reformulation markers across disciplines shows that most of them in the Lithuanian data set occur in the humanities, closely followed by the field of technology (f/10 000 = 9.4 vs. 8.8) with no statistically significant difference between the two science areas (LL = +0.76). This is an interesting finding as one might expect discourse trends in technological texts to be closer to the biomedical field rather than that of the humanities. A closer look at the data sources, however, reveals that over 50% of data in the technology texts of CorALit come from textbooks, and 14% of all reformulation markers found in the technology sub-corpus occur in textbooks (cf. only 6% of reformulation markers in textbooks in medical discourse and as little as 0.5% in the humanities, but there is also a lower percentage of textbooks included into the sub-corpora representing these two science fields). In his description of the textbook genre Hyland (2009: 118) notes the concern of textbook authors with "what the audience can be expected to know and what needs to be spelt out". Indeed, it is quite natural to have a higher density of explanations, definitions, and reformulation in a genre the ultimate goal of which is to explain and teach. It seems, therefore, that the use of the markers under analysis could be influenced not only by the language, but also by the genre of the texts.

In the English data set the quantitative distribution of reformulation markers under study is slightly different. Even though the highest number of reformulation markers is observed in the humanities, just like in the Lithuanian sub-corpus, the lowest number of them occurs in technology texts, with medical sub-corpus constituting an intermediate case. The normalised frequency of the use of reformulation markers in English medical and humanities texts seems to be quite close (3.7 vs. 4.1), however, the log likelihood test shows a statistically significant difference in their use (LL = -21.49) with the humanities scholars employing a higher number of reformulation markers in their texts.

Hyland's study (2007) showed that it was the hard sciences texts that relied heavier on reformulations in comparison to the soft fields. However, Hyland's study focused on research articles in individual disciplines. The humanities field in his study is represented by Philosophy texts with the reformulation markers' frequency of 15.3 per 10 000 words, and Applied Linguistics, where the observed density of reformulation markers is 21.0 per 10 000 words. We can see that there is a substantial difference in the distribution of reformulation markers between these two disciplines alone. The sub-corpus of academic humanities in the COCA con-

sists of research articles from various journals representing different disciplines of the humanities: *Music Educators Journal*, *Journal of Popular Culture*, *African Arts*, *Scandinavian Studies*, *Journal of Speech, Language and Hearing Research*, *Style*, *General Music Today*, etc. It could have been the case that this variety of disciplines with their specific writing practices could have influenced the overall frequency rate of the use of reformulation markers in the humanities field, so discipline could also be an important factor in the reformulation marker distribution.

The next step of the quantitative analysis was to look at how different reformulation markers are distributed in academic texts in both languages, in other words, what the variety of their use is (Table 3).

Table 3. The distribution of different reformulation markers in Lithuanian and English academic texts

Sub-corpus	*t.y.* 'i.e.' / *i.e.*	*tai yra* 'that is' / *that is to say*	*kitaip tariant* 'in other words' / *in other words*	*būtent* 'namely' / *namely*
CorALit (LT)	82%	3%	13%	2%
COCA (EN)	43%	34%	16%	7%

It is immediately obvious that the abbreviated forms are the most popular markers of reformulation in both languages: in Lithuanian it is *t.y.* which is the abbreviation of *tai yra* 'that is', and in English it is *i.e.* from the Latin *id est* 'that is'. The Lithuanian data displays a slight preference for one more marker of reformulation, which is *kitaip tariant* 'in other words', whereas in English the full form of reformulation *that is* is also very popular, leaving *in other words* as the third most frequent option to reformulate in the English language academic texts. The same order of occurrence of the three most frequent reformulation markers is also observed in the multidisciplinary corpus of research articles written by English native speakers in Murillo's (2018) study. These are only general trends, however, as the way these markers distribute in different science fields is not identical (Table 4).

Even though there is a distinct preference for the abbreviated forms mentioned above in most of the science fields, Table 4 shows that it is the humanities that display most variation in both languages. Scholars in the English humanities more readily use the full form *that is* rather than the Latin abbreviation *i.e. In other words* is also most frequently observed in the humanities field, whereas in the sciences there is heavier reliance on *i.e.*

Table 4.   The distribution of different reformulation markers in Lithuanian and English academic texts in different science fields

Sub-corpus	*t.y.* 'i.e.' / *i.e.*	*tai yra* 'that is' / *that is to say*	*kitaip tariant* 'in other words' / *in other words*	*būtent* 'namely' / *namely*
(Bio)medical LT	95%	4%	1%	-
Medical EN	71%	15%	8%	6%
Humanities LT	69%	5%	23%	3%
Humanities EN	24%	44%	23%	9%
Technology LT	90%	2%	6%	2%
Technology EN	49%	34%	11%	6%

In Lithuanian academic discourse the difference between the humanities and the other two science fields is even more readily observed, with *t.y.* 'i.e.' constituting 69% of the whole use of reformulation in the humanities, the remaining 31% split between *kitaip tariant* 'in other words', *tai yra* 'that is' and such rarer forms as *būtent* 'namely'. In his cross-disciplinary study of code glosses, Hyland also observes that "there is greater variation in the use of devices in the soft fields" (2007: 273). This is perhaps an unsurprising finding as texts in the humanities could be expected to have more variety of expression than those in the hard fields.

4.2. *Functions of reformulation markers*

For the functional analysis two reformulation markers have been chosen: 1) the abbreviated forms *t.y.* 'i.e.' and *i.e.* in all three science fields and both languages; 2) *kitaip tariant* 'in other words' / *in other words* in the three science fields in English and only in the humanities in Lithuanian as this reformulation marker has been very scarce in the texts of biomedical and technological academic discourse. As demonstrated in Table 4 above, around 90% of the reformulation marker use in biomedical and technological texts in CorALit is taken by *t.y.* 'i.e.', which means that other markers are not frequent at all.

The online randomiser resource was used at www.randomizer.org/ to generate 100 random sentence numbers for each sub-corpus in the two languages for *t.y.* 'i.e.' and *i.e.* and 50 random sentence numbers for *kitaip tariant* 'in other words' and *in other words* in the selected sub-corpora. Since the Lithuanian sub-corpus, unlike the one from the COCA, consists not only of texts from academic journals, but includes other genres, cases of use of reformulation markers in other genres were not included into the analysis. These randomly selected cases of reformulation markers in academic journals were analysed in context to determine the functions

they performed.

As mentioned in the Data and methods section, following Murillo (2012) the functions of reformulation markers were divided into three broad categories: 1) interpretation of explicit content; 2) explication of conceptual knowledge; 3) interpretation of implicit content.

Tables 5–6 below show the distribution of the above mentioned three broad functional categories of reformulation markers in both languages and the three science fields.

Table 5. Main functions performed by reformulation markers *t.y.* 'i.e.' and *i.e.* in the sub-corpora under study

Sub-corpus	Explicit (explanations, specifications)	Conceptual (definitions, denominations)	Implicit (conclusions, calculations)
(Bio)medical LT	58%	23%	19%
Medical EN	59%	31%	10%
Humanities LT	61%	31%	8%
Humanities EN	69%	27%	4%
Technology LT	63%	26%	11%
Technology EN	62%	27%	11%

Table 6. Main functions performed by reformulation markers *kitaip tariant* 'in other words' and *in other words* in the sub-corpora under study

Sub-corpus	Explicit (explanations, specifications)	Conceptual (definitions, denominations)	Implicit (conclusions, calculations)
Medical EN	62%	0%	38%
Humanities LT	74%	2%	24%
Humanities EN	35%	0%	65%
Technology EN	54%	0%	46%

The results in Table 5 show that the functions of the reformulation markers *t.y.* 'i.e.' and *i.e.* distribute quite similarly in the two languages. The interpretation of explicit content is the prototypical function taking up more than 50% of the analysed cases of use in every discipline both in Lithuanian and in English. Providing explanations of the arguments and specifications of the propositional content are

important aspects of preparing academic text for the reception of the audience to avoid misinterpretation or lack of information:

(11) *When washing or feeding the child, the figure is also washed and fed, **i.e.**, offered plates of food, which is expected to disappear.* (H-COCA)

(12) *Although we wish that our final survey had allowed a better determination of how course characteristics affected the student experience of the project, our sample size (**i.e.**, eight courses) was too small to allow any statistical insight.* (T-COCA)

(13) *All St. Patrick Company employees are on salary. You manually enter all changes (**i.e.**, new employees, salary increases, etc.) into the payroll database monthly.* (T-COCA)

In (11) the author describes rituals with wooden twin figures performed by the Temne people in the northern Sierra Leone. As the figure is wooden and cannot in fact be fed, the author explains what its feeding means. In some cases reformulation introduced by *i.e.* or its Lithuanian equivalent could act as a reminder of the already presented proposition material. The sentence in (12) occurs towards the end of the article in the discussion of the results, and *i.e.* here introduces a reminder of the sample size that was already presented earlier, apparently to remind the reader of the important information. Sometimes the use of reformulation markers resembles the use of exemplification markers, as it seems that the author(s) signal possible options rather than provide a definite, exhaustive list. A case in point is presented in (13), where the impression of exemplification is strengthened by *etc.* at the end of parentheses.

The second functional category, that is the explication of conceptual knowledge, is also distributed fairly similarly with regard to the use of *t.y.* 'i.e.' and *i.e.* in both languages and the three science fields. Here definition or denomination is employed by the authors of the articles to explain different concepts and terms:

(14) *Some fungi produce mycotoxins (**i.e.**, toxic fungal metabolites).* (M-COCA)

(15) *In film, sharpness, although allied to resolution, is not the same thing since it relates to the transition between a light tone and an adjoining darker tone (**i.e.**, acutance), and this does have a direct equivalent in digital images.* (T-COCA)

So far as the third functional category is concerned, *t.y.* 'i.e.' and *i.e.* are not typically used to signal conclusions or summary of the arguments. There were some

such cases, particularly in the humanities, but these were not that frequent, especially in the English texts. As we will see later, scholars prefer *in other words* to introduce conclusions. In (bio)medical sciences and technology texts, when *t.y.* 'i.e.' and *i.e.* performed the function of interpretation of implicit content, this typically involved introduction of mathematical calculations, as in (16):

(16) *Kiekvieną grupę sudarė 2 pogrupiai po 100 viščiukų, **t. y.** iš viso po 200 paukščių kiekvienoje grupėje.* (B-CorALit)
'Every group was composed of two subgroups, 100 chicks each, i.e. there were 200 birds in each group.'

Summing up the patterns of employment of *t.y.* 'i.e.' and *i.e.*, we can say that their functional profile is very similar, and the use in different disciplines is also quite uniform. There is one clearly observable difference between the two languages though: the English *i.e.* occurs predominantly in parentheses in all three science fields. For example, within the English humanities, out of 1165 cases of use of *i.e.* 63% occur in parentheses, whereas in the medical texts the use of *i.e.* in parentheses constitutes as many as 86%. In contrast, in Lithuanian medical texts only 4% of the use of the abbreviated form *t.y.* 'i.e.' is found in parentheses. The situation is similar in the other two science fields as well. The same tendency of the use of *i.e.* in parentheses has been noted by Murillo (2012: 82) in her dataset of Business Management research articles in English. As Murillo explains, parenthetical reformulations do not disrupt the linearity of discourse, and that may be the reason why they are so frequent in English.

The functional analysis of *kitaip tariant* 'in other words' and *in other words* (Table 6 above) again shows some similarities between the different science fields in English and English and Lithuanian humanities. These reformulation markers are usually never used to introduce the explication of conceptual knowledge (i.e. definition and denomination). Typically, they signal that a rewording or specification of the argument will follow (i.e. interpretation of explicit content function), or especially in the case of English humanities it is quite usual to find a summary of the main arguments introduced by *in other words* (65% of all functional uses of the marker):

(17) ***In other words**, though scientists may have begun to explain the mysterious links between objects and bodies through recourse to a mechanistic model, the dynamic forces powering   this model still seemed fantastic and sublime.* (H-COCA)

Example (17) presents a prototypical summarising statement introduced by *in other words*; it is the last sentence of a passage where the main arguments are outlined and it occurs at the end of the article (the penultimate passage).

Likewise, in the discourse of technology and medicine in English *in other words* is typically used to summarise rather than to present mathematical operations; the latter function seems to be much more typical for *i.e.* Even in cases when technically there are numbers present, the argument still seems to have a flavour of a concluding, generalising remark. Consider (18):

(18) *Because of multiple-unit ownership, an estimated 130 million refrigerators can be found in the 114 million households existing in the United States today. **In other words**,   roughly one in five U.S. households possesses more than one refrigerator.* (T-COCA)

A general observation concerning the functional profiles of *t.y.* 'i.e.' and *i.e.* as well as *kitaip tariant* 'in other words' / *in other words* is that their use in biomedical and technological texts displays a noticeable similarity both between the languages and across the science fields. This result echoes the observation by Duszak (1997: 11) on a more likely homogeneity of experimental sciences in comparison to the humanities and social sciences which tend to display considerable diversity of expression: "if experimental sciences are prone to show more similarities in textualization patterns, writings in the humanities and social sciences evidence more prominent variation".

Indeed, texts in the humanities displayed quite interesting differences in the use of reformulation markers. Perhaps the most obvious difference could be observed within the interpretation of explicit concept functional category. In the examples coming from more discursive disciplines, such as, for example, Arts, Religion, Philosophy, the explanations introduced by reformulation markers frequently did not lead to more specific arguments, but remained on an abstract level:

(19) *Mano antroji tezė: meniškai vaizduojama mirtis apimdama etinius ir estetinius aspektus išplečia mūsų egzistencinės kūrybos horizontą, **t.y.** turi pozityvų tikrovės statusą. **Kitaip ją galima suformuluoti taip**: meno kūrinio supratimas skleidžiasi kaip estetinių ir etinių nuostatų polifonija ir suponuoja egzistencinę kūrybą. **Kitaip tariant**, pasyvumas suprantant meno kūrinį reikalauja aktyvumo kuriant egzistencinį projektą.* (H-CorALit)
      'My second thesis: combining ethical and esthetic principles, death in art expands the horizon of our existential creativity, i.e. it has the status of a positive reality. It is possible to formulate it in other way: the perception

of a piece of art unfolds like a polyphony of esthetic and existential principles and presupposes existential creativity. In other words, the passivity in understanding a piece of art requires action in the creation of an existential project.'

(20) *Tai dieviškos skeveldros ir kita prasme: pasaulis steigiamas meilės aktu, o įsimylėti reiškia individualizuoti ženklu. **Kitaip tariant**, meilė čia reiškia ir bendražmogišką ar net dievišką patirtį, **t. y.** esmę, ir individualų jausmą kaip pasaulio žiūros tašką* (H-CorALit)
'It is a godly splinter in a different sense as well: the world is established by an act of love, and to fall in love means to be marked by a sign. In other words, love here means   interpersonal or even godly experience, i.e. the essence, and individual feeling as a way to see the world.'

Example (19) contains three reformulation markers following one another, however, explanations or reformulations that are offered by the author still stay on quite an abstract level. The same can be observed in (20) where we have two reformulation markers, but a similar level of abstraction. It is natural that a philosophical or religious discussion of death or love is not straightforward or very tangible. Thus, the reformulations shown in examples (19–20) resemble the train of thought of the author himself/herself, as if s/he tries to think aloud, grasp the thread of the argument and lead the reader alongside through the labyrinth of ideas.

However, in some texts of the humanities sub-corpora of both languages we have examples with a very clear progression from the more abstract notion to a more specific one. Several reformulation markers following each other closely can also be observed here:

(21) *In addition to things like "task chunking" (**i.e.**, doing like things together -- **in other words**, answering all my e-mail at once or running a bunch of errands in one trip), I look for short cuts to make the most of my writing time.* (H-COCA)

(22) *However, unless students are convinced of the importance, or value, of this form of intellectual activity, transfer is still unlikely to happen. **In other words**, if students are to   transfer critical thinking skills acquired in music classes to nonmusical contexts (**i.e.**, art and mathematics classes and the business world), they will need motivation, instruction, and practice to that end.* (H-COCA)

In (21) the author uses *i.e.* to provide a definition of what *task chunking* is, and immediately expands the definition by giving specific examples to illustrate it.

Likewise, in (22), which comes from *Music Educators Journal*, the author expands the idea of the transfer of critical thinking to nonmusical contexts and for the sake of clarity goes deeper into the nonmusical contexts explaining what these are. The examples containing several reformulations one after another reveal the formulation efforts of the writer (cf. Gülich & Kotschi 1983)[2].

If we look at the composition of the humanities sub-corpora in both CorALit and COCA we can see that many texts come from such disciplines as Arts, Philosophy, Literary Criticism, Religious Studies, Philology. Scientific arguments in many of these disciplines are quite abstract, relying on argumentation itself rather than on empirical tangible data. However, samples randomly selected for the functional analysis of reformulation markers also come from Education, Musicology, Linguistics, etc. which frequently deal with more concrete issues at hand. That could explain at times conceptually different use of reformulation markers in the analysed texts of the two languages despite the fact that they belong to the same science field of the humanities.

5. Concluding remarks

This exploratory study investigated the use of reformulation markers in two research cultures, Lithuanian and English, and in academic texts from three broad science fields, the humanities, technology, and (bio)medicine. The study shows that reformulation markers are more frequent in Lithuanian academic texts than in English. This result could point towards differences in rhetorical principles typical to the writing style of the two cultures, though more studies are necessary to confirm or reject this hypothesis.

In terms of the disciplinary comparison, both Lithuanian and English scholars in the humanities employ reformulation markers to the biggest extent. It is true that Lithuanian texts in the field of technology have a similar frequency of reformulation markers as the ones in the humanities, however, this is due to the differences in the generic composition of the data sets. The technology sub-corpus of CorALit includes more data from textbooks than the two other science fields. The humanities also display the widest variety of reformulation markers as well as the most diverse contexts in which they occur. The use of reformulation markers in technology and biomedical fields is much more uniform. This is in line with other scholars' observations on the structural and textual homogeneity that the hard sciences tend to dis-

---

[2] I would like to thank an anonymous reviewer for pointing out this aspect to me.

play; the soft fields, in contrast, tend to be more varied and diverse in their textual expression. It also highlights the role different disciplinary practices can play within the same science field with regard to the rhetorical weaving of the argument.

So far as the functional profile of reformulation markers *t.y.* 'i.e.' and *i.e.* as well as *kitaip tariant* 'in other words' / *in other words* is concerned, interpretation of the explicit meaning is the most frequent function of these markers in Lithuanian and English academic discourse in most of the science fields. Other functions (introducing conclusions, mathematical operations, definitions, etc.) are also present, but they are distributed differently for different markers.

These results are important in that they shed more light on what is typical to particular science fields and particular research cultures. In order to succeed as scientific writers, researchers have to be aware of the discourse features that are customary and acceptable in their discipline. The knowledge of the rhetorical conventions and writing style is crucial for showing oneself as an insider in the discipline. Finally, extending research on the use of metadiscourse elements in languages other than English helps to highlight universal patterns of research writing as well as specific features typical only to some cultural communities.

Further avenues for the study of reformulation markers in academic discourse could include their frequency and functional analysis across different genres (research articles, textbooks, monographs) and separate disciplines rather than broad science fields. It would also be important (and interesting) to investigate how learners use reformulation markers in comparison to professional researchers. Another area of future research on reformulation markers could be their use in translated texts. Finally, it has to be acknowledged that the present study only looked at a fixed set of reformulation markers; there are other, less obvious means to signal reformulation in academic discourse, such as parentheses, for example. These also have to be investigated in more detail in the future.

Data sources

COCA: Corpus of Contemporary American English. Available at: https://corpus.byu.edu/coca/

CorALit: Lietuvių mokslo kalbos tekstynas [Corpus Academicum Lithuanicum]. Available at: www.coralit.lt

References

Aijmer, Karin. 2007. The meaning and functions of the Swedish discourse marker alltså: evidence from translation corpora. *Catalan Journal of Linguistics* 6. 31–59.

Atkinson, Dwight. 2004. Contrasting rhetorics/contrasting cultures: why contrastive rhetoric needs a better conceptualization of culture. *Journal of English for Academic Purposes* 3(4). 277–289. http://dx.doi.org/10.1016/j.jeap.2004.07.002

Biber, Douglas & Johansson, Stig & Leech, Geoffrey & Conrad, Susan & Finegan, Edward. 1999. *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.

Blakemore, Diane. 1993. The relevance of reformulations. *Language and Literature* 2(2). 101–120. http://dx.doi.org/10.1177/096394709300200202

Bondi, Marina. 2008. Emphatics in academic discourse: integrating corpus and discourse tools in the study of cross-disciplinary variation. In Ädel, Annelie & Reppen, Randy (eds.), *Corpora and discourse: the challenges of different settings*, 31–55. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.31.04bon

Crismore, Avon & Markkanen, Raija & Steffensen, Margaret S. 1993. Metadiscourse in persuasive writing. A study of texts written by American and Finnish university students. *Written Communication* 10(1). 39–71.

Cuenca, Maria-Josep. 2003. Two ways to reformulate: a contrastive analysis of reformulation markers. *Journal of Pragmatics* 35(7). 1069–1093. http://dx.doi.org/10.1016/S0378-2166(03)00004-3

Duszak, Anna. 1997. Cross–cultural academic communication: a discourse community view. In Duszak, Anna (ed.), *Culture and styles of academic discourse*, 11–39. Berlin: Mouton de Gruyter. http://dx.doi.org/10.1515/9783110821048.11

Fløttum, Kjersti & Dahl, Trine & Kinn, Torodd. 2006. *Academic voices: across languages and disciplines*. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/pbns.148

Gülich, Elisabeth & Kotschi, Thomas. 1983. Les marqueurs de reformulation paraphrastique. *Cahiers de Linguistique Française* 5. 15–81.

Halliday, M. A. K. 1973. *Explorations in the functions of language.* London: Edward Arnold.

Halliday, M. A. K. & Hasan, Ruqaiya. 1976. *Cohesion in English.* London: Longman. https://doi.org/10.4324/9781315836010

Hyland, Ken. 2001. Definitely a possible explanation: epistemic modality in academic argument. In Gotti, Maurizio & Dosena, Marina (eds.), *Modality in specialized texts*, 291–310. Frankfurt: Peter Lang.

Hyland, Ken. 2005. *Metadiscourse: exploring interaction in writing*. London: Continuum.

Hyland, Ken. 2007. Applying a gloss. Exemplifying and reformulating in academic discourse. *Applied Linguistics* 28(2). 266–285. http://dx.doi.org/10.1093/applin/amm011

Hyland, Ken. 2008. Persuasion, interaction and the construction of knowledge: representing self and others in research writing. *International Journal of English Studies* 8(2). 1–23.

Hyland, Ken. 2009. *Academic discourse. English in a global context.* London: Continuum.

Hyland, Ken & Sancho Guinda, Carmen (eds.). 2012. *Stance and voice in written academic genres.* Basingstoke: Palgrave Macmillan. http://dx.doi.org/10.1057/9781137030825

Mazzi, Davide. 2011. "In other words, …": a corpus-based study of reformulation in judicial discourse. *Hermes – Journal of Language and Communication Studies* 46. 11–24. http://dx.doi.org/10.7146/hjlcb.v24i46.97361

Mur-Dueñas, Pilar. 2007. 'I/we focus on …': a cross-cultural analysis of self-mentions in business management research articles. *Journal of English for Academic Purposes* 6(2). 143–162.

Mur-Dueñas, Pilar. 2011. An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics* 43. 3068–3079. http://dx.doi.org/10.1016/j.pragma.2011.05.002

Mur-Dueñas, Pilar & Šinkūnienė, Jolanta. 2016. Self-reference in research articles across Europe and Asia: a review of studies. *Brno Studies in English* 42(1). 71–92. http://dx.doi.org/10.5817/BSE2016-1-4

Murillo, Silvia. 2012. The use of reformulation markers in business management research articles: an intercultural analysis. *International Journal of Corpus Linguistics* 17(1). 69–90. http://dx.doi.org/10.1075/ijcl.17.1.03mur

Murillo, Silvia. 2018. Not the same, but how different? Comparing the use of reformulation markers in ELF and in ENL research articles. In Mur-Dueñas, Pilar & Šinkūnienė, Jolanta (eds.), *Intercultural perspectives on research writing*, 237–253. Amsterdam: John Benjamins.

Sanderson, Tamsin. 2008. Interaction, identity and culture in academic writing: the case of German, British and American academics in the humanities. In Ädel, Annelie & Reppen, Randy (eds.), *Corpora and discourse: the challenges of different settings*, 57–92. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.31.05san

Sperber, Dan &Wilson, Deirdre. 1986. *Relevance.* Blackwell: Oxford.

Šinkūnienė, Jolanta. 2011. *Autoriaus pozicijos švelninimas rašytiniame moksliniame diskurse: gretinamasis tyrimas* [Hedging in written academic discourse: a cross-linguistic and cross-disciplinary study]. Unpublished doctoral dissertation. Vilnius: Vilnius University.

Šinkūnienė, Jolanta. 2018. The power of English: I and we in Lithuanian, Lithuanian English and British English research writing. In Mur-Dueñas, Pilar & Šinkūnienė, Jolanta (eds.), *Intercultural perspectives on research writing*, 59–79. Amsterdam: John Benjamins.

Usonienė, Aurelija & Butėnas, Linas & Ryvitytė, Birutė & Šinkūnienė, Jolanta & Jasionytė, Erika & Juozapavičius, Algimantas. 2011. Corpus Academicum Lithuanicum. In

Vetulani, Zygmunt (ed.), *Human language technology. Challenges for computer science and linguistics. Lecture notes in computer science*, 412–422. Berlin: Springer.

Vande Kopple, William J. 1985. Some exploratory discourse on metadiscourse. *College Composition and Communication* 36(1). 82–93.

Vaskó, Ildiko & Fretheim, Thorstein. 1997. Some central pragmatic functions of the Norwegian particles altså and nemlig. In Swan, Toril & Westvik, Olaf Jansen (eds.), *Modality in the Germanic languages*, 233–292. Berlin: Mouton de Gruyter. http://dx.doi.org/10.1515/9783110889932.233

Vassileva, Irena. 2001. Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes* 20. 83–102. http://dx.doi.org/10.1016/S0889-4906(99)00029-0

Author's address:

Vilnius University; Department of English Philology
Universiteto str. 3, Vilnius, LT – 01513, Lithuania
E-mail: jolanta.sinkuniene@flf.vu.lt

OBJASNIDBENI KONEKTORI U AKADEMSKOM DISKURSU: MEĐUJEZIČNO I MEĐUDISCIPLINARNO ISTRAŽIVANJE

Na temelju kontrastivne analize te uporabom kvantitativnog i kvalitativnog metodološkog pristupa, u radu se istražuju distribucija i funkcije objasnidbenih konektora u korpusu tekstova iz humanističkih, medicinskih i tehničkih znanosti u engleskom i litvanskom jeziku. Engleski primjeri prikupljeni su iz dijela (potkorpusa) Korpusa suvremenog američkog engleskog jezika (COCA) koji čine znanstveni tekstovi, dok je za analizu litvanskog jezika korišten specijalizirani sinkronijski korpus pisanog akademskog litvanskog jezika (Corpus Academicum Lithuanicum). Analiza pokazuje da se u oba jezika objasnidbeni konektori najčešće pojavljuju u tekstovima pisaca iz humanističkih znanosti. Rezultati također ukazuju na to da u odnosu na pisce tekstova iz tehničkih i medicinskih znanosti, pisci u humanističkim znanostima koriste raznovrsnije konektore za ostvarivanje više različitih funkcija u tekstu. Neovisno o jeziku i znanstvenom području, najčešća funkcija objasnidbenih konektora je interpretacija eksplicitnog sadržaja. Rad naglašava važnost znanstvenog područja i žanra u distribuciji i uporabi objasnidbenih konektora.

Ključne riječi: objasnidbeni konektori; akademski diskurs; korpusno utemeljen; engleski; litvanski.