



10.2478/topling-2022-0009

Glossing an argument: Reformulation and exemplification in L2 Master's theses

Tereza Guziurová*
University of Ostrava, Czech Republic

Abstract

Following a large body of research on metadiscourse in academic writing, this paper explores one feature of textual metadiscourse, code glosses, in English L2 academic texts written by Czech university students. The study draws on Hyland's metadiscourse model (2005), which characterizes code glosses as devices that elaborate propositional meanings by rephrasing or explaining what has been said. Thus, they can help readers understand the writer's intended meaning or contribute to the formation of persuasive arguments. The corpus consists of 48 English L2 Master's theses representing three disciplines – linguistics, literature and English language teaching (ELT) methodology, totalling almost 950,000 words. The results are compared with professional writing represented by English L1 research articles from the same disciplines. The findings reveal differences in the frequency and functions of several code glosses, as novice writers are shown to overuse certain devices. The findings also indicate cross-disciplinary variation, as reformulation and exemplification proved to be much more prominent in linguistics and methodology than in literary studies.

Key words

metadiscourse, code glosses, reformulation, exemplification, Master's theses, research articles

1. Introduction

For over 30 years, metadiscourse has been used as an important analytical framework for investigating writer-reader interaction in academic genres. In his recent appraisal of the concept, Hyland (2017, p. 16) argues that metadiscourse is perhaps now “one of the most commonly employed methods for approaching specialist written texts”. The ongoing interest in the term is undoubtedly connected with a more general trend of investigating interpersonal aspects of academic discourse. Mauranen (2012, p. 45) called this trend the “interpersonal turn” in the early 1990s, when written academic discourse began to be systematically analysed in terms of interaction between writers and readers, leading to the development of such concepts as metadiscourse (Guziurová, 2018).

This trend is corroborated by two bibliometric studies following changes in English for Academic Purposes (EAP) and applied linguistics research. Hyland and Jiang (2021) followed research trends in EAP over the last 40 years based on analysis of 40 journals that most often publish EAP-related articles. They found that during that period some topics experienced a significant rise in interest, e.g. topics dealing with *identity*, *genre*, *interaction* and *discipline* (Hyland and Jiang, 2021). This suggests a growing interest in academic literacy and the ways academics argue their claims and communicate with readers within discourse communities (*ibid.*).

* Address for correspondence: Tereza Guziurová, Department of English Language Teaching, Faculty of Education, University of Ostrava, Fráni Šrámka 3, 709 00 Ostrava, Czech Republic. E-mail: tereza.guziurova@osu.cz

Lei and Liu (2019) conducted a similar bibliometric study, tracking research trends in applied linguistics from 2005 to 2016. Generally, they conclude that topics which witnessed significantly increased interest are especially those related to sociocultural and identity issues and issues of new technology, while some others (particularly those concerning traditional phonological and grammatical issues) have become ever less popular (Lei and Liu, 2019, p. 555). Substantial attention devoted to topics such as *identity*, *interaction*, *genre* and the relation between texts and contexts might explain why metadiscourse remains so popular in applied linguistics research. Despite the fact that the term itself is quite fuzzy and there are different approaches to the concept, metadiscourse continues to function as an analytical framework for investigating academic genres across different languages and disciplines.

This study draws on Hyland's conception of metadiscourse, which is defined as "the linguistic resources that the writer uses to explicitly organize the text, to refer to its parts or to other texts, to express his stance towards it and to interact with the reader" (Hyland and Tse, 2004, p. 157). Understanding metadiscourse primarily as writer-reader interaction, Hyland distinguishes two main types of metadiscourse – interactive and interactional. **Interactive** resources are text-oriented and concern "the ways writers signal the arrangement of their texts based on their appreciation of the reader's likely knowledge and understandings" (Hyland, 2005, pp. 43–44). The organization of the text is thus perceived as guiding readers through the content by anticipating their likely reactions and needs, which also depends on what the author knows about his/her audience. **Interactional** resources concern ways writers comment on and evaluate content, and ways they express commitment to propositions and interact with their readers. This type of metadiscourse includes the categories of "stance" and "engagement" (Hyland, 2005). While interactional resources, including hedges, boosters, evidentials (i.e. citation practices) and engagement markers, have been heavily studied in academic writing (e.g. McGrath and Kuteeva, 2012; Hyland, 2014; Hu and Cao, 2015; Hyland and Jiang, 2016; Dontcheva-Navratilova, 2021, etc.), interactive resources have attracted less attention (probably with the exception of transition markers, which comprise conjunctions and adverbial phrases signalling relations in the text).

This study focuses on one feature of interactive resources, code glosses, which are used to explain, define or clarify a writer's intended meaning and thus contribute to the formation of persuasive arguments. The study aims to explore how novice writers use code glosses in the genre of English L2 Master's theses and compare this with professional writing represented by English L1 research articles from the same disciplines. Specifically, the main aims of the study are:

- (1) to compare the frequency of occurrence of code glosses in L2 Master's theses written by Czech university students with their occurrence in L1 research articles in three "soft" disciplines, namely linguistics, literature and English language teaching (ELT) methodology;
- (2) to find out whether there are cross-disciplinary differences in the use of code glosses in L2 Master's theses;
- (3) to identify which functions selected code glosses convey in the two corpora.

The analysis thus encompasses two dimensions, learner writing versus professional writing, and English L1 versus L2 discourse.

The study is part of a larger research project focusing on writer-reader interaction in English-medium academic discourse written by Czech university students. Generally, the aims of the project are to identify intercultural differences in the use of metadiscourse in English-medium academic writing between Czech and British novice writers and to compare the learner discourse to expert Anglophone academic discourse and explain contextual and linguistic factors motivating divergences across the different levels of expertise. For these purposes, three corpora will be contrasted, the third (British students' academic writing) of which is currently in preparation.

This article focuses only on the two corpora available so far and works with two variables, intercultural variation and degree of expertise. Even though this analysis provides only partial results of the larger project, it takes into consideration both variables in the interpretation of results, based on previous research into the topic. It also takes account of similarities between the two genres in terms of organization, since what has been referred to as the "traditional thesis" usually follows the typical "IMRD" structure (introduction – methods – results – discussion), complemented only by a literature review chapter (Paltridge and Starfield, 2020, p. 88). This multidimensional comparison aims to

contribute to a large body of research on L2 learner corpora as it characterizes English-medium Master's theses written by Czech university students.

2. Background

Code glosses help readers understand the text by rephrasing or explaining what has been said and they are introduced by phrases such as *that is*, *in other words*, *for example*, etc. A detailed explanation of the term will be provided in section 3. Code glosses have been investigated within academic discourse in different genres and disciplines. Hyland's early studies of metadiscourse, for example, showed that code glosses were a common feature of popular science articles and undergraduate textbooks, where they worked to clarify unfamiliar terms or concepts (Hyland, 2005, p. 97). My previous analysis of metadiscourse also emphasized the role of code glosses in university textbooks: they were almost twice as numerous as in research articles from the same discipline, reducing the cognitive load of propositional content for novices in the field (Guziurová, 2018).

Cross-disciplinary investigation of code glosses has shown that they fulfil different functions in different disciplines. Specifically, natural sciences rely more heavily on reformulation to make observations and interpretations more specific, while social sciences and humanities are "more explicitly interpretative, using reformulation to draw implications" for a potentially more diverse audience (Hyland, 2007, p. 284). Writers in "soft" disciplines generally need to place greater emphasis on situating their research and explaining their claims, as research in "soft" fields does not always occur within the established theoretical paradigm (Hyland, 2007). As a result, writers cannot always assume that the background to a research problem, appropriate methods and terminology are agreed on by all readers, and they have to clarify the context and define concepts to show how they are used in their particular study.

As part of metadiscourse, code glosses have also been studied cross-linguistically, e.g. in Chinese (Mu et al., 2015) and Spanish (Mur-Dueñas, 2011). Both studies use Hyland's (2005) model of metadiscourse (albeit slightly modified) and investigate research articles within a single discipline. Mu et al. (2015) found that there were generally more metadiscourse features in English RAs (written by English native speakers or authors affiliated with English-speaking institutions) than in Chinese RAs (written by Chinese native speakers). In particular, English writers used twice as many code glosses as Chinese writers in their texts. Similarly, Mur-Dueñas (2011) discovered that metadiscourse features were significantly more numerous in international English-medium RAs written by authors from American universities than in Spanish RAs written by Spanish scholars. Some of her findings suggest that "American-based scholars more closely guide the readers through their arguments in their RAs especially by means of logical markers and code glosses" (Mur-Dueñas, 2011, p. 3071).

Moreover, features which are functionally equivalent to code glosses have been studied independently of metadiscourse, e.g. Murillo (2012) compared the use of reformulation markers (which are part of code glosses) in L1 and L2 research articles. The results showed that reformulation markers in business management RAs occurred far more frequently in articles written in L1 English than in those written in L1 Spanish. L2 English articles written by Spanish academics were closer to L1 English articles suggesting that "Spanish academics tend to adapt to English conventions when writing their research in English for international publications" (Murillo, 2012, p. 73).

Given the prevalence of the metadiscourse framework, both textual and interpersonal features have been heavily investigated in learner corpora.¹ An influential model of metadiscourse was provided by Ädel (2006), who used the Jakobsonian model of language functions as a basis for her definition of metadiscourse. She compared argumentative essays written by Swedish EFL university students and L1 university students (both British and American). Her study revealed a general pattern of considerable overuse of metadiscourse by Swedish learners as they used twice as many metadiscourse features as American students, who, in turn, used more markers than British students (Ädel, 2006, p. 189).

Metadiscourse in postgraduate writing was also investigated by Hyland (2004b). The analysis of L2 Master's and doctoral dissertations has shown that hedges and transitions were the most common devices overall. In terms of disciplinary differences, students in "soft" (social sciences and humanities)

¹ For a detailed account of corpus-based analyses of metadiscourse in learner corpora, see Neff-van Aertselaer (2015).

disciplines have been found to employ more metadiscourse, especially interactional (“textual”) features. This also applies to code glosses, which Hyland found to be most frequent in applied linguistics dissertations (Hyland, 2004b, p. 146).

Although a number of studies have already investigated discourse markers in learner texts by comparing novice English L1 writers and novice L2 writers with different mother tongue backgrounds (e.g. Granger and Tyson, 1996; Altenberg and Tapper, 1998), the main focus has remained on cohesive devices, i.e. text-specific features that can reveal patterns of text organization (Neff-van Aertselaer, 2015) or, to a lesser extent, on specific pragmatic discourse features, especially endophoric markers (Bunton, 1999) and exemplifying signals (Paquot, 2008). Within this framework, the present study aims to investigate a less explored feature of textual metadiscourse, i.e. code glosses, in L2 Master’s theses written by Czech university students in three disciplines – linguistics, literature and English language teaching (ELT) methodology.

3. Code glosses

Discourse features expressing reformulation or clarification of the writer’s meaning are present in a certain form in most conceptions of metadiscourse, whether it is a broad approach including stance and writer-reader interaction (e.g. Crismore, 1989; Hyland, 2005) or a narrow approach focusing on text organization and elements referring to the text itself (e.g. Mauranen, 1993; Ädel, 2006).² Hyland’s (2005) model of metadiscourse has been selected for this study since it is complex and has been frequently applied in recent genre-based studies (e.g. Del Saz Rubio, 2011; McGrath and Kuteeva, 2012; Kawase, 2015; Mu et al., 2015, etc.). As previously stated, this analysis is part of a larger research project focusing on writer-reader interaction in L2 Master’s theses and Hyland’s framework allows for investigation of both textual and interpersonal features in this genre. Moreover, it builds on previous taxonomies and reorganizes the categories of metadiscourse in a systematic way.

In Hyland’s metadiscourse model (2005), code glosses are characterized as follows:

Code glosses supply additional information, by rephrasing, explaining or elaborating what has been said, to ensure the reader is able to recover the writer’s intended meaning. They reflect the writer’s predictions about the reader’s knowledge-base and are introduced by phrases such as *this is called*, *in other words*, *that is*, *this can be defined as*, *for example*, etc. (Hyland, 2005, p. 52)

Two broad subfunctions of code glosses are distinguished: reformulation and exemplification (Hyland, 2007). Reformulation promotes textual cohesion and facilitates discursive progression as it provides “a retrogressive interpretation of the previous utterance and allows speakers to explain, rephrase, reconsider, summarize or even distance themselves from it” (Dal Negro and Fiorentini, 2014, p. 95). Two types of reformulation, paraphrastic and non-paraphrastic (Bondi, 2020), are distinguished in the literature of pragmatics. The former is based on a semantic equivalence between the two members, which are thus on the same hierarchical level. In the latter type, a new formulation is produced, resulting in an explanation, summary, conclusion, etc. (Dal Negro and Fiorentini, 2014). However, in practice there is often no clear-cut distinction between the two types, so some authors reject the distinction altogether (e.g. Del Saz Rubio and Fraser, 2003).

Hyland’s conception of code glosses (2007), which this study draws on, does not seem to consider any reformulation as purely paraphrastic. He argues that while the two parts might have the same propositional content, the pragmatic meaning would always be different. Furthermore, it should be noted that Hyland’s conception of reformulation is broad. It covers not only traditional reformulation markers which are grammaticalized (such as *i.e.*, *that is*, *namely*, *in other words*), but also various lexical means used to rephrase or explain what has been said (*this means*, *this can be defined as*, etc.) (Guziurová, 2020).

The second broad function of code glosses is exemplification. According to Hyland (2007, p. 270), it is a “process through which meaning is clarified or supported by a second unit which illustrates the first by citing an example”. Generally, exemplification in academic texts can help support the writer’s claims and thus make arguments more persuasive. It can also help explain terms, concepts

² For a detailed discussion of a “broad” (integrative) approach and a “narrow” (non-integrative) approach, see Ädel (2006).

or definitions, allowing the reader to grasp the writer's intended meaning. Hyland (2007, p. 279) distinguishes three main ways in which examples work in academic writing: 1) by offering an instance of a general category; 2) by providing a parallel or similar case; and 3) by giving a precept or a rule. The detailed functions of reformulation and exemplification will be discussed in the next section.

4. Data and methodology

4.1 Corpus and method

The research is carried out on a specialized corpus of English-medium Master's theses written by Czech university students (MT corpus). The corpus consists of 48 theses representing three disciplines – linguistics (16 theses), literature (16 theses) and ELT methodology (16 theses), totalling almost 950,000 words altogether. The authors are postgraduate students majoring in English language and literature at Masaryk University in Brno, Czech Republic, and their L1 is Czech. In order to represent a good standard of student writing, only theses graded “A” (“Excellent” / “Merit”) were included in the corpus. All the theses were submitted between 2010 and 2018.

In order to explore the typical metadiscourse features of the MT corpus, a reference corpus was compiled consisting of published research articles (RA corpus). The corpus consists of 36 research articles and is fully comparable in terms of disciplines, representing linguistics (12 RAs), literature (12 RAs) and methodology (12 RAs). All the papers were written by Anglophone authors based at British or American universities. Although it cannot be guaranteed that all of the authors are native speakers of English, based on their names, affiliations and CVs (if available) it is reasonable to assume that their L1 is English. We selected only well-established international journals which are indexed in the Web of Science database (Social Sciences Citation Index, 2020 or Arts & Humanities Citation Index, 2020). Each discipline was represented by three prestigious journals. All the articles were published between 2010 and 2018 and are single-authored. The reference corpus is smaller than the MT corpus, totalling only 242,439 words. However, it is believed that being comparable in terms of disciplines and time span, it can show whether there are differences in the use of code glosses between expert and novice writers. In addition, the results can be compared with previous studies on the use of metadiscourse in research articles. Table 1 shows the composition and size of both corpora.

Table 1. Composition of the MT corpus and the RA corpus

	Linguistics		Literature		ELT methodology		Total	
	No. of texts	No. of words	No. of texts	No. of words	No. of texts	No. of words	No. of texts	No. of words
MT corpus	16	260,582	16	327,764	16	359,146	48	947,492
RA corpus	12	76,700	12	82,878	12	82,861	36	242,439

Prior to the analysis, all texts were processed. Bibliographic references, abstracts, tables and figures were omitted, as well as citations in the text (block quotes as well as in-text citations longer than four words). The analysis was thus carried out on the author's own text, which is the subject of interest in metadiscourse research. The data in Table 1 refer to the size of the corpora after they have been processed, so the number of words in the RA corpus might seem small.

The texts were first searched for specific features which could potentially act as code glosses, based on the list from Hyland's study (2007) and my own previous research on metadiscourse (Guziurová, 2020) using the AntConc concordancer. Unlike Hyland (2007), I focused only on explicit textual devices, so that punctuation markers, such as parentheses, were not regarded as metadiscourse markers for the purposes of this study. The search terms for reformulation included *i.e.*, *that is*, *in other words*, *namely*, *mean(s)*, *called*, *defined as*, *refer* to* (*refers*, *referred*, *referring*), *known as*, *termed*, *specifically*, *more precisely*, *or x* and *put it*. The search terms for exemplification included *such as*, *e.g.*, *for example* and *for instance*. Then all the cases were examined in context to ensure that they functioned as code glosses. For example, the word *mean* is polysemous, but only the sense “to have a particular meaning (to signify)” was considered reflexive. Meanings like “to intend” were disregarded.

The results were normalized to 100,000 words to allow comparison across the two corpora. Then log-likelihood tests were performed to determine differences of statistical significance. All statistical tests were performed on raw data using the online calculator developed by Paul Rayson (Rayson, 2008). Currently, there is a debate in applied linguistics (and other fields) about the limitations of inferential statistics and p-values in the research process (Brezina, 2018; Rayson, 2008), and there is an increasing tendency to calculate effect size measures which “express the practical importance of the effect observed in the corpus or corpora” (Brezina, 2018, p. 14). For this reason, the Log Ratio (binary log of the ratio of relative frequencies) was also used in some instances to determine the size of the difference for the two corpora (Hardie, 2014). After the quantitative analysis, the functions of the most frequent code glosses were determined by means of close reading of the texts.

4.2 Analytical framework

This study primarily follows Hyland’s (2007) typology of the discourse functions of code glosses, which distinguishes two broad subcategories: reformulation and exemplification. Reformulation is “a discourse function whereby the second unit is a restatement or elaboration of the first in different words, to present it from a different point of view and to reinforce the message” (Hyland, 2007, p. 269). According to Hyland, reformulation can work in two ways, either to expand the reader’s understanding of the text (by means of explanation or implication), or to reduce the scope of interpretation (by means of paraphrase or specification). Hyland’s typology of functions is given in Figure 1.

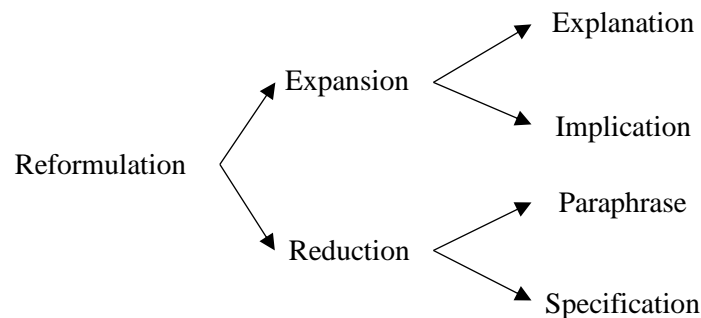


Figure 1. Discourse functions of reformulation (Hyland, 2007, p. 274)

The model appears to be informed by previous studies of reformulation, and particularly by Del Saz Rubio and Fraser (2003), who distinguish “expansion” and “compression” in their typology. As mentioned above, Hyland’s conception does not consider any reformulation as purely paraphrastic, since it always expands the original or reduces it. What seems problematic is Hyland’s understanding of expansion: in his account of metadiscourse it is not the material (or the information) that is expanded, but the reader’s understanding of it. This is connected with his basic premise that metadiscourse is distinct from propositional content (see Hyland, 2005). However, in practice, it is difficult to decide whether reformulation functions to expand the idea itself or just the sense in which the writer intends it to be understood, and the distinction between propositional content and metadiscourse is thus quite blurred.

This study considers all the cases of reformulation which rephrase or explain what has been said to help the reader recover the writer’s intended meaning as code glosses. The reformulations are signalled by reformulation markers as explicitness is an important criterion of metadiscourse. As Hyland puts it, metadiscourse refers to “aspects of a text which explicitly organize a discourse or the writer’s stance towards either its content or the reader” (2005, p. 14). However, cases which generally characterize the concept as part of an argument (main proposition) are not considered metadiscursive, for example “Flemish is a Germanic language and is very closely related to English language” (MT_METH_06). This might be considered as a definition with a classifying function by Triki (2019), but is not regarded as a code gloss in this study.

The individual functions of reformulation are characterized below (Hyland, 2007, pp. 274-276). The examples are taken from the MT corpus.

- A) Explanation – a clarification which elaborates the meaning of a preceding unit to facilitate its comprehension by providing a gloss or a definition. It includes cases where a technical term is explained (1), or where a term is provided for a concept already expressed (2).
- (1) This is an interlingual error, i.e. an error influenced by L1 of the learners. (MT_METH_11)
 - (2) Rods and cones are not distributed evenly on the retina - there are more cones around the centre of the retina, called fovea, and much fewer cones outside the fovea. (MT_LING_05)
- B) Implication – serves to draw a conclusion or to sum up the main import of the previous unit. It is often signalled by *in other words*.
- (3) Observing the students’ reactions showed that they were all really active and motivated. They were looking forward to finding out something new about these places. In other words, the texts used definitely corresponded with the learners’ interests. (MT_METH_16)
- C) Paraphrase – restating an idea in different words to provide a summary. It is typically signalled by the forms *that is*, *in other words* or *put another way*.
- (4) There were fourteen participants out of thirty (i.e. 46.7%) who were able to imitate the pronunciation without being instructed. (MT_METH_03)
- D) Specification – providing further details of the statement in order to constrain how the reader might interpret it. Typical signals are *specifically* or *namely*.
- (5) Indispensably, by establishing the ‘we’ identity, Cherry does not refer only to himself and to the co-host Ron MacLean, but he acknowledges the presence of the audience, more specifically Canadian audience, which is crucial in negotiating collective Canadian consciousness. (MT_LING_03)

The explanation function as defined by Hyland is broad since it covers both definition and denomination processes (which are usually treated separately, see e.g. Murillo, 2018). Furthermore, it is questionable whether the denomination processes, represented by example (2), can be considered as a form of expansion. Rather, they seem to restrict the scope of interpretation by giving the exact term to the phenomenon (see Triki, 2019). Consequently, I decided to disregard Hyland’s distinction of expansion and reduction processes and I followed a combination of Hyland’s (2007) framework and Murillo’s (2012; 2018; 2019) classification of the functions of reformulation markers that proved functional in different kinds of texts (journalistic, academic etc.) (Murillo, 2018, p. 240).

5. Results and discussion

5.1 An overview of code glosses in the two corpora

At first, the two corpora were contrasted in terms of the frequency of occurrence of code glosses. The results of the quantitative analysis are given in Table 2. Owing to the different sizes of the corpora, all the numbers were normalized to 100,000 words.

Table 2. General frequency of code glosses in the MT corpus and the RA corpus

Function	MT corpus			RA corpus		
	Total no.	Frequency per 100,000 words	% of total CG	Total no.	Frequency per 100,000 words	% of total CG
Reformulation	1545	163.0	44.0	328	135.0	40.8
Exemplification	1968	207.6	56.0	475	195.4	59.2
Code glosses (TOTAL)	3513	370.6	100	803	330.4	100

CG = code glosses

The general frequency of code glosses is higher in the Master’s theses corpus, with 370.6 cases per 100,000 words, and the difference is statistically significant ($G^2=8.51$; $p<0.01$)³. In both corpora, exemplification predominates over reformulation, which corresponds to Hyland’s findings that exemplification plays a significant role in “soft” disciplines as it represents a “heavier rhetorical investment in contextualization” (Hyland, 2007, p. 272). In Hyland’s study of 240 research articles, the ratio of the two functions varied depending on the disciplines, with 61% of code glosses in the hard sciences signalling reformulation and 61% of code glosses in the social sciences and humanities signalling exemplification (ibid.).

The overall frequency of exemplification markers was similar in the MT corpus and in the RA corpus (with 207.6 and 195.4 occurrences per 100,000 words, respectively), and there is no statistical difference between the corpora ($G^2=1.32$; $p>0.05$). It was thus in the function of reformulation that the two corpora varied the most, as there were 163.0 occurrences per 100,000 words in the MT corpus and 135.0 in the RA corpus ($G^2=9.82$; $p<0.01$). The next section will look at these differences in more detail, considering different forms and functions that reformulation markers fulfil.

5.2 Reformulation – forms and functions

In order to account for quantitative differences in the use of reformulation markers, let us first consider which markers were used in the two corpora. Table 3 shows a distribution of specific code glosses expressing reformulation in the Master’s theses and research articles.

Table 3. Frequency of specific reformulation markers in the MT corpus and the RA corpus

Marker	MT corpus		RA corpus	
	Total/Frequency per 100,000 words	% of reformulation markers	Total/Frequency per 100,000 words	% of reformulation markers
i.e.	483/50.9	31.3	57/23.5	17.4
that is	103/10.9	6.6	65/26.7	19.8
in other words	228/24.1	14.7	34/14.0	10.4
namely	131/13.8	8.5	18/7.4	5.5
mean(s)	143/15.1	9.2	31/12.8	9.5
called	117/12.3	7.6	20/8.2	6.1
defined as	46/4.9	3.0	12/4.9	3.7
refer* to	91/9.6	5.9	26/10.7	7.9
known as	20/2.1	1.3	9/3.7	2.7
termed	6/0.6	0.4	10/4.1	3.0
specifically	68/7.2	4.4	27/11.1	8.2

³ Log-likelihood tests were performed to determine whether the results were statistically significant. If the p-value was <0.05 (the threshold level usually set in linguistics), the results were regarded as statistically significant (following e.g. Dontcheva-Navratilova, 2021).

more precisely	16/1.7	1.0	0/0	0
or x	75/7.9	4.9	18/7.4	5.5
put it	18/1.9	1.2	1/0.4	0.3
Total	1545/163.0	100	328/135.0	100

The data show that the most common reformulation marker in the MT corpus is the abbreviation *i.e.*, followed by *in other words*, *mean(s)*, *namely* and *called*. In the reference corpus, the order is slightly different, the most frequent reformulation marker being *that is*, followed by *i.e.*, *in other words*, *mean(s)* and *specifically*. It is worth noting that *i.e.* accounts for 31.3% of all reformulation markers in the MT corpus, which makes it by far the most popular means of signalling equivalence among students. The most frequent device in the RA corpus, *that is*, accounts only for 19.8% of all reformulation markers, suggesting a greater variation.

If we compare the results with the use of code glosses in English as a lingua franca (ELF) academic writing (Guziurová, 2020, p. 44), we can find certain similarities. ELF writers also preferred the abbreviation *i.e.* when reformulating their statements (accounting for 27.2% of all reformulation markers), while native speakers used more varied devices (*i.e.* accounting for 17.1% only) (ibid.). Czech students seemed to have preferred *i.e.* as a simple and grammatically fixed marker which does not pose any problems in production and/or understanding. It is interesting to note that when studying ELF research articles written by researchers from 10 different L1 backgrounds (the SciELF corpus), Murillo (2018) found that *i.e.* was the most popular marker among Czech writers (who used it much more frequently than other L1 writers). This suggests that both Czech novice and expert writers rely on *i.e.* heavily when reformulating.

The greatest quantitative differences between the corpora were found in the use of the markers *i.e.* ($G^2=37.36$; $p<0.0001$) and *that is* ($G^2=29.52$; $p<0.0001$). The Log Ratio for the marker *i.e.* is 1.12, which means that it is two times more common in the Master’s theses than in the research articles. On the contrary, *that is* occurs twice more frequently in the research articles (Log Ratio=1.30). A detailed analysis of the functions allows us to account for these differences. Table 4 shows the distribution of functions in the two corpora, using Hyland’s framework as introduced in section 4.

Table 4. Functions of the markers *i.e.* and *that is* in the two corpora (raw frequencies and percentages)

Function	MT corpus				RA corpus			
	<i>i.e.</i>	<i>that is</i>	Total	%	<i>i.e.</i>	<i>that is</i>	Total	%
Explanation	125	37	162	27.8	13	23	36	29.5
Implication	18	11	29	5.0	5	11	16	13.2
Paraphrase	135	33	168	28.8	24	24	48	39.3
Specification	202	22	224	38.4	15	7	22	18.0
Total	480	103	583	100	57	65	122	100

Overall, the most frequent function in the MT corpus was specification (38.4%). Novice writers used reformulation markers *i.e.* and *that is* quite frequently to specify their statements in all three disciplines. Indeed, specification was the most frequent function of the marker *i.e.* in the Master’s theses; it allowed the authors to “highlight the specification and simultaneously include it within the scope of the original formulation” (Hyland, 2007, p. 276) and the readers to interpret the message correctly (6).

- (6) Newspeak evolved on the basis of invention of new words, fundamental reduction of vocabulary, i.e. elimination of undesired words, and removal of unorthodox meanings. (MT_LIT_03)

One of the types of specification is identification, in which the reformulation marker helps to identify the referent of the previous phrase (Murillo, 2018, p. 240). In the following example (7), “respondents” are identified as “college students” in a coreference process. The “college students” have been

mentioned earlier in the section when the author discusses a previous research study; however, he/she reminds the readers who the “respondents” are to ensure they interpret the message correctly.

- (7) Horwitz et al. (1986) worked with the same sample of respondents, *i.e.* college students, and as such, it was an appropriate source to be used in the development of the FLCAS. (MT_METH_09)

The specification was also realized by enumerating (8). In this case the cataphoric element (“five phases”) is followed by an enumeration of phases, which has a discourse organizing function (Murillo, 2018).

- (8) The new course has been redesigned in the Moodle software, but this time according to the rules of the ADDIE model, following its five phases, *i.e.* analysis, design, development, implementation, and evaluation, thus exploring the ADDIE method and its possible utilisation in online language learning. (MT_METH_04)

Both enumeration and identification were much more frequent in the Master’s theses than in the research articles where specification seems to have been signalled also lexically, by the marker *specifically*.

The second most frequent function in the MT corpus is paraphrase followed by an explanation. As mentioned above, the function of explanation is very broad and several subfunctions were found in the corpus. Firstly, the writers used *i.e.* and *that is* to define concepts. This function seems to be particularly important in soft disciplines in which the terminology is not completely unified and thus the writers aim to show in which sense the concepts are used for the purposes of their study. The definitions sometimes included a citation or a reference to a source (9).

- (9) The keystone of their theory is a technical term *face*, *i.e.* “a public self-image of a person that every [adult] member [of a society] wants to claim for himself” (Brown and Levinson 1987: 61). (MT_LING_11)

In fact, definitions (but also general characterizations of concepts) were quite frequent in the Master’s theses since their authors wanted to demonstrate that they had understood them. This may reflect one of the main communicative goals of the genre, *i.e.* showing that the author is familiar with the basic terminology of the discipline. However, in order to show their understanding, students sometimes tend to overexplain, which is apparent in the following example (10), where language use is being elaborated on, maybe unnecessarily, as an “interaction between individuals (oral or written)”, followed by a number of examples of general aspects influencing it.

- (10) As the language is being used, *i.e.* an interaction between individuals (oral or written), it is happening at a particular time and place and at the same time it is being influenced by many other aspects like ways of thinking, behaviour, values, emotions, beliefs, ideologies, perspectives, as well as things like dress, gestures and symbols among others. (MT_LING_09)

The second subfunction is referred to as a denomination by Murillo (2018), as the writer provides a term or a designation for a concept already explained (11).

- (11) My intention is to analyze a corpus of sentences implementing the Pr-Scale in fiction narrative, focusing mainly on the syntactic and semantic characteristics of the verbs that are able to express existence or appearance on the scene (*i.e.* Presentation verbs). (MT_LING_2)

Explanation also includes cases where writers elaborate the meaning of a preceding unit in order to clarify the meaning and ensure the correct interpretation. In the following example the author explains the concept “full error analysis” (12).

- (12) Unlike some of the researchers whose studies I presented in Chapter 1, I conducted a full error analysis, i.e. all deviations from the norm were annotated and no particular linguistic feature was preferred at the expense of another. (MT_METH_11)

Paraphrasing, or restating an idea in different words to provide a summary, as Hyland (2007) puts it, was heavily used in both corpora. In the Master's theses paraphrases often took the form of quantifications (13) or clarifications which narrowed down the scope of interpretation (14).

- (13) As regards the syntactic characteristics of the verbs, 88 out of the total of 121 (i.e. 72.7%) manifest an intransitive use. (MT_LING_02)

- (14) The language is often informal since the characters find themselves in natural settings among their friends, peers, work colleagues, i.e. people they feel comfortable with. (MT_LING_11)

Paraphrasing was prominent in research articles in which the authors sought to clarify their statements to get the message across and also to support their arguments (15). In Example (16), the writer adds the information needed for the reader to understand the message correctly.

- (15) Teacher efficacy is cyclical in nature. That is, a proficient performance creates a new mastery experience, which provides new information that is processed to shape future efficacy beliefs. (RA_METH_05)

- (16) If the numeral + classifier construction in Mandarin is indeed taking on the function of the English indefinite article, Mandarin learners may map this form to the indefinite article in the L2 in contexts where the numeral sounds natural in Mandarin (i.e., with neutral and noninferable referents). (RA_METH_10)

The third reformulation marker which was used significantly more frequently in the corpus of Master's theses was *in other words* ($G^2=9.83$; $p<0.01$). This marker typically introduces implication in research articles (Murillo, 2018), but it can also perform other functions, namely explanation and paraphrase. While expert writers often used this marker to make implications and draw conclusions, novice writers seemed to express other, less complex, operations, such as paraphrase. In fact, a considerable number of cases are paraphrases of quotes or examples since the writers wanted to show they understood them and were able to comment on them properly (17). As an ability to work with academic sources and to understand information in context is a crucial part of the evaluation of theses, students commented on the citations quite often. Furthermore, it should be noted that the marker *in other words* was heavily overused in one literary thesis, which included 77 occurrences (out of a total 228).

- (17) Following the definition of instruction presented above, Merrill et al. (1996) defined it as “a technology for the development of learning experiences and environments which promote the acquisition of specific knowledge and skill by students” (p. 6). In other words, it is a carefully guided process of creating courses and activities that help students maximize their learning. (MT_METH_14).

An interesting case of reformulation marker is *namely*. It was almost twice as numerous in theses as it was in research articles and the difference is statistically significant ($G^2=7.14$; $p<0.01$). Similarly, the marker proved to be overused by French learner writers (Granger and Tyson, 1996) and it was also popular among ELF authors of research articles in comparison with L1 English writers (Guziurová, 2020). Furthermore, it was misused by several students, who either could not properly integrate it into the sentence (as in example 18), or used it in the wrong context, so the resulting sentence is semantically incorrect (19). Such cases were not attested in the L1 research articles.

- (18) Whether a metaphorical expression has been employed deliberately or non-deliberately is namely a matter of the communicative value of metaphor. (MT_LING_14)

- (19) Because the present perfect tense does not exist in L1 of the learners, namely it may be translated as the past or present tenses depending on the context, this error may be tagged as an interlingual one, i.e. an error that is the result of L1 influence (Ellis & Barkhuizen, 2008). (MT_METH_11)

The results concerning reformulation are consistent with those of Ädel (2006), who found that Swedish EFL learners overuse metadiscourse in comparison with native speakers. Czech learner writers specifically overused the marker *i.e.*, since they used it twice as frequently as expert writers. This may reflect one of the main communicative purposes of the Master’s thesis genre, as the authors defined concepts quite often in order to demonstrate that they had understood them. However, *i.e.* also fulfilled other functions, most notably specification, which was often realized by enumeration or identification in the theses. One student used *i.e.* systematically in the contexts where examples were provided, so he/she probably mistook *i.e.* for *e.g.* Taken together with the use of *namely* mentioned above, it appears that even some advanced learners of English have not fully mastered the use of reformulation markers. The next section will discuss general trends in exemplification in both corpora.

5.3 Exemplification in Master’s theses and research articles

Exemplification is an important part of argumentation in academic writing since writers can support their arguments by providing specific examples to general statements. Both expert and novice writers recognized the rhetorical potential of examples since they used them with almost equal frequency, as can be seen in Table 5.

Table 5. Frequency of specific exemplification markers in the MT corpus and the RA corpus

Marker	MT corpus		RA corpus	
	Total/Frequency per 100,000 words	% of exemplification markers	Total/Frequency per 100,000 words	% of exemplification markers
such as	890/93.9	45.2	160/65.8	33.7
e.g.	499/52.6	25.4	134/55.1	28.2
for example	368/38.8	18.7	136/56.0	28.6
for instance	211/22.3	10.7	45/18.5	9.5
Total	1968/207.6	100	475/195.4	100

Four expressions were chosen for analysis (*for example*, *e.g.*, *for instance*, *such as*) as these were found by previous studies (e.g. Hyland, 2007, p. 278) to be the most frequent exemplification markers in social sciences. As Table 5 illustrates, the overall frequency of exemplification markers is similar in both corpora. Moreover, the proportional distribution of individual markers is comparable, with *such as* being the most popular device with both novice and expert writers. This is consistent with other studies of exemplification in academic writing, which have found that *such as* is favoured in a number of disciplines (e.g. Hyland, 2007; Triki, 2021). Nevertheless, when we compare the frequency of the top marker *such as* in the two corpora, we find that it occurs significantly more often in Master’s theses ($G^2=18.31$; $p<0.0001$). On the other hand, *for example* occurs more frequently in research articles ($G^2=12.64$; $p<0.001$).

Several trends can be observed in the data. First, the use of the abbreviation *e.g.* in the research articles was restricted exclusively to parentheses, which corresponds with Triki’s (2021) latest findings on exemplification. All but one of 134 cases occurred in parentheses, typically introducing a nominal group or a reference to another source (20). Expert writers used *e.g.* with references quite frequently to support their arguments or situate their research within a discipline.

- (20) This development has been described as one of the most significant developments in im/politeness research (e.g. Mitchell and Haugh, 2015) and a central tenet of the discursive approach (e.g. Terkourafi, 2005; van der Bom and Mills, 2015). (RA_LING_05)

The use of *e.g.* was more varied in the Master's theses as only 66% of cases occurred in parentheses and references to sources were much less frequent. This suggests that students were not fully aware of this rhetorical convention, i.e. use of *e.g.* to introduce background information in academic writing (see Biber et al., 1999).

Second, several studies have pointed out that the marker *for instance* has been significantly overused by advanced learners of English (Altenberg and Tapper, 1998; Paquot, 2008). This was not the case in the MT corpus, even though the marker was used with slightly greater frequency by novice writers. Nevertheless, it was the least frequent marker of exemplification in both corpora, and in the research articles it was used by only 9 authors (out of 36), thus suggesting that this pattern may be related to the author's personal style (see Biber et al., 1999, p. 890).

The last interesting trend concerns differences among disciplines. While *such as* was heavily used throughout all the disciplines, *e.g.* was restricted to linguistics and methodology, especially in the RA corpus. In fact, literature articles included only one example of *e.g.* The reason can be found in the house rules of the journals included in the corpus, one of which explicitly states that most Latin abbreviations in text are discouraged.⁴ This shows the importance of genre and disciplinary conventions in metadiscourse. Disciplinary variation in the use of code glosses will be discussed in detail in the next section.

5.4 Code glosses across disciplines

It is now a well-established fact that disciplinary variation in the use of metadiscourse is high (see e.g. Hyland, 2005; 2007). Sometimes referred to as academic *cultures* or *tribes* (Hyland, 2012), academic disciplines have their own norms and practices which constitute separate cultures. According to Hyland (2004a, p. 8), these cultures “differ along social and cognitive dimensions, offering contrasts not only in their fields of knowledge, but in their aims, social behaviours, power relations, political interests, ways of talking and structures of argument”. Within each culture individuals acquire specialized discourse competencies that allow them to participate as group members (Hyland, 2004a), and it is these competencies that novice writers aspire to learn.

The analysis of code glosses, although representing only one feature of metadiscourse, also revealed disciplinary differences. Table 6 shows the distribution of code glosses across three disciplines – linguistics, literature and methodology.

Table 6. Distribution of code glosses across disciplines

		LING	LIT	METH	Total
		Total/ Freq. per 100,000 w	Total/ Freq. per 100,000 w	Total/ Freq. per 100,000 w	Total/ Freq. per 100,000 w
MT corpus	Reformulation	571/218.7	432/131.7	542/151.0	1545/163.0
	Exemplification	586/224.5	465/141.7	917/255.4	1968/207.6
	CG (Total)	1157/443.2	897/273.4	1459/406.4	3513/370.6
RA corpus	Reformulation	137/178.0	63/75.9	128/154.2	328/135.0
	Exemplification	189/245.4	76/91.6	210/253.0	475/195.4
	CG (Total)	326/423.4	139/167.5	338/407.2	803/330.4

CG = code glosses

The data indicate that the overall distribution of code glosses across the disciplines is similar in the two genres, with the highest frequency of markers in linguistics and the lowest in literature. The quantitative differences between linguistics and methodology are quite small, but there are significant differences between the two disciplines and literary studies. This suggests that literature, which traditionally belongs to the humanities, might have different rhetorical conventions and modes of argumentation from the other two disciplines. Concerning the MT corpus specifically, the linguistics

⁴ <https://ecf.humanities.mcmaster.ca/submissions/> [Accessed 20 December 2021].

and methodology theses used mostly quantitative methodology (e.g. questionnaires and experiments in ELT methodology and analyses of language data which are then quantitatively processed in linguistics). In contrast, the literature theses used critical and interpretative methods of inquiry, i.e. mostly a qualitative design. The same applies to the RA corpus, in which linguistics and methodology articles followed an IMRD organization pattern, which is typical of a quantitative research design (Gray, 2015).

The findings are in line with Cao and Hu (2014), who discovered that quantitative research articles in their corpus of 120 papers from three soft disciplines used markedly more reformulators than qualitative research articles. In their view, this can be explained by “the greater importance that quantitative research attaches to precision and specification of knowledge claims and their scope of generalization” (Cao and Hu, 2014, p. 27). They believe that abstract, technical knowledge in quantitative research often requires elaboration to facilitate comprehension (ibid.). A detailed analysis of disciplinary differences in the MT corpus and the RA corpus is given below.

As far as reformulation is concerned, the findings have revealed that there were differences in the specific markers across disciplines (Table 7). In linguistics, Czech novice writers overwhelmingly favoured *i.e.* when reformulating their statements, while expert writers preferred *that is*, followed by *i.e.* In methodology, *i.e.* was the top marker in both genres. In literature, expert writers relied on *that is*, followed by the lexical marker *mean**, but the data are quite small to draw any conclusions.

Table 7. Distribution of reformulation markers across disciplines

Marker	REFORMULATION					
	LINGUISTICS		LITERATURE		METHODOLOGY	
	MTs	RAs	MTs	RAs	MTs	RAs
	Total/ Freq. per 100,000 words		Total/Freq. per 100,000 words		Total/ Freq. per 100,000 words	
<i>i.e.</i>	231/88.5	16/20.8	54/16.5	1/1.2	198/55.2	40/48.2
<i>that is</i>	9/3.5	43/55.8	67/20.4	12/14.5	27/7.5	10/12.0
<i>in other words</i>	51/19.5	9/11.7	114/34.8	6/7.2	63/17.5	19/22.9
<i>namely</i>	43/16.5	9/11.7	44/13.4	5/6.0	44/12.3	4/4.8
<i>mean*</i>	61/23.4	14/18.2	29/8.8	10/12.0	53/14.8	7/8.4
<i>called</i>	55/21.1	6/7.8	14/4.3	6/7.2	48/13.4	8/9.6
<i>defined as</i>	21/8.0	6/7.8	6/1.8	1/1.2	19/5.3	5/6.0
<i>refer* to</i>	35/13.4	12/15.6	18/5.5	3/3.6	38/10.6	11/13.2
<i>known as</i>	3/1.1	0/0	7/2.1	3/3.6	10/2.8	6/7.2
<i>termed</i>	0/0	3/3.9	3/0.9	4/4.8	3/0.8	3/3.6
<i>specifically</i>	28/10.7	14/18.2	24/7.3	3/3.6	16/4.5	10/12.0
<i>more precisely</i>	5/1.9	0/0	5/1.5	0/0	6/1.7	0/0
<i>or x</i>	28/10.7	5/6.5	33/10.1	8/9.6	14/3.9	5/6.0
<i>put it</i>	1/0.4	0/0	14/4.3	1/1.2	3/0.8	0/0
Total	571/218.7	137/178.0	432/131.7	63/75.9	542/151.0	128/154.2

The most frequent reformulation marker in literature theses was *in other words*. As mentioned earlier, although this result was partly due to the overuse of this marker in one thesis, it is worth stating that this marker was mainly employed to introduce interpretations of literary passages and examples (21).

- (21) Furthermore, in a different section its narrative voice confides his secret to the reader: “A.J. is an agent like me, but for whom or for what no one has ever been able to discover. It is rumored that he represents a trust of giant insects from another galaxy . . . I believe he is on the Factualist side (which I also represent)” (122-23). In other words, the reader is tempted to believe the discourse’s main narrator is an agent for the Factualists, [...]. (MT_LIT_10)

Results also showed that code glosses explicitly concerning terminology, i.e. *defined as*, *refer to*, *known as*, *called*, were widely used in linguistics and methodology to carry out multiple communicative

functions related to defining the main theoretical concepts, comparing conceptions formulated by different authors or discussing methodological frameworks. In literature, these lexical markers were scarce (in both genres), and they appeared to be mainly employed to intertextually refer to concepts and labels taken from other literary texts, such as in the following example, where the writer provides a definition of the term “doublethink” from the novel *1984*.

(22) The Newspeak term doublethink refers to such a status of mind when a person is able to accept and identify oneself with two contradictory propositions without resulting in doubts or even impeachment of credibility of the propositions: [...]. (MT_LIT_03)

Another aspect which stood out in literature was the absence of abbreviated markers *i.e.* and *e.g.* in the research articles (1 occurrence of each in the whole corpus). As noted in the previous section, guidelines of the literary journals discouraged the use of Latin abbreviations in texts, so writers used only full forms *that is* and *for example*. Unlike expert writers, students used abbreviated forms repeatedly, but overall these were lower in number than full forms.

If we again examine the functions of the reformulation markers *i.e.* and *that is* across disciplines, we find the following pattern (see Table 8). Overall, the most frequent function in linguistics proved to be paraphrase (34.1%), closely followed by specification (32.4%). In literature, the most frequent function was also paraphrase (38.1%) although it should be noted again that literature authors reformulated their statements much less frequently than authors in the other two disciplines (and the expert writers even less so).

Table 8. Functions of the markers *i.e.* and *that is* across disciplines (raw frequencies and percentages)

Function	LING				LIT				MET			
	<i>i.e.</i>	<i>that is</i>	Total	%	<i>i.e.</i>	<i>that is</i>	Total	%	<i>i.e.</i>	<i>that is</i>	Total	%
Explanation	59	22	81	27.1	17	19	36	26.9	62	19	81	29.8
Implication	11	8	19	6.4	3	9	12	8.9	9	5	14	5.1
Paraphrase	87	15	102	34.1	19	32	51	38.1	53	10	63	23.2
Specification	90	7	97	32.4	16	19	35	26.1	111	3	114	41.9
Total	247	52	299	100	55	79	134	100	235	37	272	100

While in linguistics, the paraphrase was typically introduced by *i.e.*, in literature it was typically introduced by *that is*. Furthermore, literature authors often used paraphrases to clarify or even translate some expressions from literary works (23), while in linguistics paraphrases often functioned as quantifications (24) or short clarifications.

(23) While the working-class Farrens have “only porridge, and too little of that” (Brontë 105) for dinner, and the child labourers in Moore’s mill are treated to “the little cans which held their coffee, and to the small baskets which contained their allowance of bread” (Brontë 46), the Moore household abounds in delicacies: “stewed pears and ... a good-sized Belgian tartine” (Brontë 48) are served for breakfast, and Hortense Moore has a quarrel with the maid over “the bouillon ... and ... choucroute” (Brontë 49), that is clear beef broth and sauerkraut. (MT_LIT_11)

(24) It has been identified in 14 sentences, i.e. 10%. (MT_LING_02)

The most frequent function of the markers *i.e.* and *that is* in methodology was specification (41.9%). As mentioned in section 5.2, specification often took the form of enumeration in which a more general term is followed by a list and thus specified (25). This function occurred repeatedly both in methodology and linguistics.

(25) The fourth chapter considers the potential sources of FLA in six broad categories, i.e. personal and interpersonal anxieties, learner beliefs about language learning, instructor beliefs about

language teaching, instructor-learner interactions, classroom procedures, and language testing. (MT_METH_09)

As far as exemplification is concerned, the most prominent marker in the theses was *such as*, irrespective of the discipline (see Table 9 below). Students in all three disciplines preferred *such as*, which typically introduces nominal groups (Triki, 2021). On the other hand, *for example* and *for instance* were found to be preferred for the introduction of clauses in the previous studies (Triki, 2021). The use of exemplification markers was more varied in research articles, as linguists favoured *for example*, literature writers *such as* and methodology writers *e.g.* However, the distribution among individual markers was more balanced, which concurs with Hyland's (2007) findings.

Table 9. Distribution of exemplification markers across disciplines

Marker	EXEMPLIFICATION					
	LINGUISTICS		LITERATURE		METHODOLOGY	
	MTs	RAs	MTs	RAs	MTs	RAs
	Total/ Freq. per 100,000 words		Total/Freq. per 100,000 words		Total/ Freq. per 100,000 words	
such as	254/97.3	55/71.4	229/69.8	35/42.2	407/113.4	70/84.3
e.g.	145/55.6	45/58.4	51/15.5	1/1.2	303/84.4	88/106.0
for example	117/44.8	63/81.8	107/32.6	27/32.5	144/40.1	46/55.4
for instance	70/26.8	26/33.8	78/23.8	13/15.7	63/17.5	6/7.2
Total	586/224.5	189/245.4	465/141.7	76/91.6	917/255.4	210/253.0

6. Conclusion

This study has investigated the use of code glosses in English-medium Master's theses written by Czech university students. The results have been compared with professional writing represented by L1 research articles from the same disciplines. The results have shown that novice writers recognise the importance of reformulation and exemplification in their argumentative practices since they use code glosses frequently. Indeed, quantitative comparisons indicate an overuse of reformulation markers in their theses. This may be partly given by the character of the genre, which requires that they demonstrate knowledge and understanding of the theories, methods and nomenclature of a given discipline. However, in comparison with the L1 writers, the Czech students overused two expressions, *i.e.* and *such as*, irrespective of discipline. This suggests a certain tendency to simplification since the students relied on simple, grammaticalized forms, which do not require much processing effort. The findings may also support the claim that while the general overuse of reformulation markers in the theses may be given by the lower level of the writers' expertise, specific markers the authors opt for (*i.e.*, *such as*) tend to be influenced by cultural differences and Czech academic writing conventions (Dontcheva-Navratilova, 2021).

As for disciplinary variation, the two genres followed a similar pattern, at least in terms of quantitative differences, showing that reformulation and exemplification are much more prominent in linguistics and methodology than in literary studies. The differences between disciplines suggest that even though researchers often distinguish two large groups – sciences on the one hand, and social sciences and humanities (SHH) on the other – we should not forget that the groups are far from homogenous, and it is often research methodology (quantitative or qualitative) which makes a difference.

There are several limitations of this study. First, the reference RA corpus is considerably smaller than the MT corpus, which makes certain comparisons difficult. Second, in order to account for all the variables, we would need the third corpus of L1 learner writing. This corpus is in preparation for investigation in the future. Despite its limitations, the present study has pointed out certain differences between academic texts written by novice and expert writers, and it has shown that a good command of metadiscourse is important for advanced learners in order to master the rhetorical conventions of a discipline. The study has pedagogical implications as it highlights the importance of writers' awareness of how discipline influences rhetorical choices in constructing academic texts. It can also

help students develop their academic writing skills by showing that they can employ a variety of resources to express reformulation and exemplification rather than limit themselves to one fixed phrase.

Acknowledgement

This study was supported by the Czech science foundation grant 21-12150S Intercultural variation in writer-reader interaction in English-medium academic discourse by Czech and Anglophone novice writers.

References

- Ädel, A., 2006. *Metadiscourse in L1 and L2 English*. Amsterdam/Philadelphia: John Benjamins.
- Altenberg, B. and Tapper, M., 1998. The use of adverbial connectors in advanced Swedish learners' written English. In: S. Granger, ed. *Learner English on computer*. London: Longman, pp. 80-93.
- Anthony, L., 2019. AntConc (Version 3.5.8) [Computer Software]. Tokyo: Waseda University. Available at: <https://www.laurenceanthony.net/software>
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finnegan, E., 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.
- Bondi, M., 2020. Academics online: Code glosses across research genres and public communication. In: G. Maurizio, S. Maci and M. Sala, eds. *Scholarly pathways. Knowledge transfer and knowledge exchange in academia*. Bern: Peter Lang, pp. 57-82.
- Brezina, V., 2018. *Statistics in corpus linguistics*. Cambridge: Cambridge University Press.
- Bunton, D., 1999. The use of higher level metatext in PhD theses. *English for Specific Purposes*, vol. 18, pp. S41-S56.
- Cao, F. and Hu, G., 2014. Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics*, vol. 66, pp. 15-31.
- Crismore, A., 1989. *Talking with readers: Metadiscourse as rhetorical act*. New York: Peter Lang.
- Dal Negro, S. and Fiorentini, I., 2014. Reformulation in bilingual speech: Italian *cioè* in German and Ladin. *Journal of Pragmatics*, vol. 74, pp. 94-108.
- Del Saz Rubio, M. M., 2011. A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of agricultural sciences. *English for Specific Purposes*, vol. 30, pp. 258-271.
- Del Saz Rubio, M. M. and Fraser, B., 2003. *Reformulation in English*. (unpublished manuscript). [Accessed 15 December 2021]. Available at: <https://people.bu.edu/bfraser/>
- Dontcheva-Navratilova, O., 2021. Engaging with the reader in research articles in English: Variation across disciplines and linguacultural backgrounds. *English for Specific Purposes*, vol. 63, pp. 18-32.
- Granger, S. and Tyson, S., 1996. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, vol. 15, no. 1, pp. 17-27.
- Gray, B., 2015. *Linguistic variation in research articles*. Amsterdam/Philadelphia: John Benjamins.
- Guziurová, T., 2018. *Metadiscourse in undergraduate textbooks and research articles in linguistics*. Ostrava: University of Ostrava.
- Guziurová, T., 2020. Discourse reflexivity in written academic English as lingua franca: Code glosses in research articles. *Discourse and Interaction*, vol. 13, no. 2, pp. 36-54.
- Hardie, A., 2014. Log ratio – an informal introduction. *cass.lancs.ac.uk* [Accessed 19 April 2022]. Available at: <http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/>
- Hu, G. and Cao, F., 2015. Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for Specific Purposes*, vol. 39, pp. 12-25.
- Hyland, K. 2004a. *Disciplinary discourses. Social interactions in academic writing*. Ann Arbor: The University of Michigan Press.
- Hyland, K. 2004b. Disciplinary interactions: metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, vol. 13, pp. 133-151.
- Hyland, K., 2005. *Metadiscourse: Exploring interaction in writing*. London and New York: Continuum.
- Hyland, K., 2007. Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, vol. 28, no. 2, pp. 266-285.
- Hyland, K., 2012. *Disciplinary identities: Individuality and community in academic writing*. Cambridge: Cambridge Applied Linguistics.

- Hyland, K., 2014. Engagement and disciplinarily: The other side of evaluation. In: G. Del Lungo Camiciotti and E. Tognini Bonelli, eds. *Academic discourse – new insights into evaluation*. Bern: Peter Lang, pp. 13-30.
- Hyland, K., 2017. Metadiscourse: What is it and where is it going? *Journal of Pragmatics*, vol. 113, pp. 16-29.
- Hyland, K. and Jiang, F. K., 2016. “We must conclude that.”: A diachronic study of academic engagement. *Journal of English for Academic Purposes*, vol. 24, pp. 29-42.
- Hyland, K. and Jiang, F. K., 2021. A bibliometric study of EAP research: Who is doing what, where and when? *Journal of English for Academic Purposes*, vol. 49, [100929].
- Hyland, K. and Tse, P., 2004. Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, vol. 25, no. 2, pp. 156-177.
- Kawase, T., 2015. Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes*, vol. 20, pp. 114-124.
- Lei, L. and Liu, D., 2019. Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics*, vol. 40, no. 3, pp. 540-561.
- Mauranen, A., 1993. *Cultural differences in academic rhetoric: a textlinguistic study*. Frankfurt am Main: Peter Lang.
- Mauranen, A., 2012. *Exploring ELF. Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.
- McGrath, L. and Kuteeva, M., 2012. Stance and engagement in pure mathematics research articles: linking discourse features to disciplinary practices. *English for Specific Purposes*, vol. 31, pp. 161-173.
- Mu, C., Zhang, L.J., Ehrich, J. and Hong, H., 2015. The use of metadiscourse for knowledge construction in Chinese and English research articles. *Journal of English for Academic Purposes*, vol. 20, pp. 135-148.
- Mur Dueñas, P., 2011. An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics*, vol. 43, no. 13, pp. 3068-3079.
- Murillo, S., 2012. The use of reformulation markers in Business Management research articles: An intercultural analysis. *International Journal of Corpus Linguistics*, vol. 17, no. 1, pp. 64-90.
- Murillo, S., 2018. Not the same, but how different? Comparing the use of reformulation markers in ELF and in ENL research articles. In: P. Mur-Dueñas and J. Šinkūnienė, eds. *Intercultural perspectives on research writing*. Amsterdam/Philadelphia: John Benjamins, pp. 237-253.
- Murillo, S., 2019. Reformulation and its markers in unpublished research articles: Some evidence on the rhetorical patterns of written academic ELF. *LFE: Revista de Lenguas para Fines Específicos*, vol. 25, no. 2, pp. 26-43.
- Neff-van Aertselaer, J., 2015. Learner corpora and discourse. In: S. Granger, G. Gilquin and F. Meunier, eds. *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press, pp. 255-280.
- Paltridge, B. and Starfield, S., 2020. *Thesis and dissertation writing in a second language*. London and New York: Routledge.
- Paquot, M., 2008. Exemplification in learner writing: A cross-linguistic perspective. In: F. Meunier and S. Granger, eds. *Phraseology in foreign language learning and teaching*. Amsterdam/Philadelphia: John Benjamins, pp. 101-119.
- Rayson, P., 2008. Log-likelihood and effect size calculator. *ucrel.lancs.ac.uk* [Accessed 19 April 2022]. Available at: <http://ucrel.lancs.ac.uk/llwizard.html>
- Triki, N., 2019. Revisiting the metadiscursive aspect of definitions in academic writing. *Journal of English for Academic Purposes*, vol. 37, pp. 104-116.
- Triki, N., 2021. Exemplification in research articles: Structural, semantic and metadiscursive properties across disciplines. *Journal of English for Academic Purposes*, vol. 54, [101039].