DISCOURSE REFLEXIVITY IN WRITTEN ACADEMIC ENGLISH AS LINGUA FRANCA: CODE GLOSSES IN RESEARCH ARTICLES

Tereza Guziurová

Abstract

There is probably no doubt that English is a lingua franca of academia nowadays. Previous research has shown that one of the tendencies observed in academic English as Lingua Franca (ELF) speech is enhanced explicitness (Mauranen 2012). A common manifestations of explicitation is discourse reflexivity (Mauranen 2017), making discourse organisation more visible and enhancing the clarity in the contexts where interlocutors do not share a common linguistic and/or cultural background. Drawing on these findings this paper aims to investigate reflexive phenomena in written academic ELF, namely in the SciELF corpus. The corpus consists of research articles written by L2 users of English, which have not undergone any professional proofreading. The paper focuses on one reflexive category, code glosses, which "supply additional information by rephrasing, explaining or elaborating what has been said" (Hyland 2005: 52). The findings are compared with the corpus of published research articles written by Anglophone writers, which has been designed on the basis of SciELF. The results reveal significant differences in the frequency and functions of several code glosses, but at the same time show the importance of discourse reflexivity in both corpora.

Keywords

discourse reflexivity, code glosses, English as a lingua franca, SciELF corpus, research articles

1   Introduction

Reflexivity is an important feature of human interaction. Often characterized as the capacity of a natural language to refer to or describe itself, it was included in Charles Hockett's list of the defining characteristics of human language (so called "design features of language"), to distinguish our unique communication potential as a species (Sinclair 2005: 164). According to Mauranen (2007: 4), discourse reflexivity is very likely to be a "discourse universal, i.e. a discourse feature found in a very large number, perhaps all, languages".

Research on discourse reflexivity, or metadiscourse, is vast, but it can also be characterized as quite fuzzy. Over the past 40 years or so, metadiscourse has been used quite frequently as a tool for characterizing various genres, especially the academic ones. Hyland (2017: 16) argues that metadiscourse is now perhaps "one of the most commonly employed methods for approaching specialist written texts".

This holds true in spite of the fact that differences between its conceptualizations are considerable and have important methodological and practical implications for research (for a detailed discussion of different approaches to metadiscourse and conceptual difficulties of the term, see Hyland 2017).

The use of discourse reflexivity has also been investigated in academic English as a Lingua Franca (ELF), especially in ELF speech. The origins of systematic research on academic ELF date back to the beginning of the 21ˢᵗ century and essential for its development was the project *English as an Academic Lingua Franc*a (ELFA project) started in 2001 at the University of Helsinki by Professor Anna Mauranen. Studies on the ELFA corpus (a 1 million word corpus of spoken academic ELF) yielded interesting results on the nature of spoken ELF, and among other things they have shown that "discourse reflexivity is central to academic discourse, and particularly relevant for academic ELF, where it can help increase clarity and explicitness among speakers from different linguistic and cultural backgrounds" (Mauranen et al. 2016: 46). Moreover, discourse reflexivity in ELF was found to have similar functions, forms and distributions across different L1 speakers (ibid.).

What has not been thoroughly investigated is written academic ELF. A written corpus of academic ELF, the WrELFA corpus, was compiled in 2015 (as part of the ELFA project), and the first studies started to examine various aspects of written ELF (e.g. Carey 2013, Vetchinnikova 2014). It should be noted, of course, that there already exists a large body of research on academic discourse in English produced by non-native speakers. Drawing on the tradition of contrastive rhetoric (Connor 1996), researchers analysed grammatical, lexical or discursive features of academic texts written by L2 speakers of English and contrasted them with texts written by native speakers even before the ELFA project (e.g. Valero-Garcés 1996, Hyland & Milton 1997, Vassileva 2001, etc.). The problem with this kind of research was that it often led to an image of a native speaker as an ideal, the only language standard which the others should seek to achieve. However, as English was gradually becoming a lingua franca of academia, the reality started to change and the majority of writers and readers of academic texts nowadays are likely to be academics from a non-English L1 background.

Within this context the analysis of the WrELFA corpus provides a unique opportunity to answer the questions about the nature of written academic ELF. This paper aims to investigate discourse reflexivity in the SciELF 2015 corpus, which is one part of the WrELFA. The SciELF consists of 150 research articles written by L2 users of English, which have not undergone professional

proofreading or checking by English native speakers. Most of the articles are final drafts of unpublished manuscripts.

The SciELF corpus has already been investigated from several linguistic viewpoints (see e.g. Rowley-Jolivet 2017, Mur Dueñas 2018, Murillo 2018 or Wu et al. 2020). Of particular importance for this paper is the study by Bondi and Borelli (2018). In their contribution, Marina Bondi and Carlotta Borelli focused on metadiscursive resources used by ELF writers, as they compared the economics component of the SciELF with a reference corpus of published articles in the same discipline. The results showed that the SciELF corpus is characterized by some prototypical metadiscursive elements, such as evidentials (*according to*), reporting verbs and nouns. The study seemed to confirm the "cooperative imperative" of ELF users resulting in the tendency to strive for greater explicitness in their communication (ibid.: 232).

Indeed, enhanced explicitness has been observed in academic ELF speech (Mauranen 2012, 2017). This tendency seems to result from the interaction between people from different linguistic and/or cultural backgrounds who cannot rely on the shared cultural context. One of the common manifestations of explicitation is discourse reflexivity (Mauranen 2017: 246), which makes discourse organisation more visible and enhances clarity by guiding readers through the discourse and making it more accessible to recipients. Discourse reflexivity thus seems to be a good research area for discovering what characterizes written ELF.

This paper focuses on one reflexive category, code glosses, which are used to help readers understand the text by rephrasing or explaining what has been said; they are introduced by phrases such as *in other words*, *that is*, *for example*, etc. The aim of the study is to investigate and compare the forms and functions of code glosses in two corpora: a corpus of written academic ELF represented by a part of the SciELF corpus, and a corpus of published research articles written by Anglophone authors. The aim is to explore the nature of written academic ELF, specifically, the results should indicate whether the tendency to enhanced explicitness is only shown in spoken ELF, or whether it is an integral part of written academic ELF as well.

The next section presents the definition of code glosses. Section 3 characterizes the two corpora and the methodology. Section 4 provides the results of the analysis and discussion. In the last section some conclusions are drawn concerning the discourse reflexivity in written ELF.

2   Code glosses

The category of code glosses is in a certain form present in most conceptions of metadiscourse, whether it is a broad approach including stance and writer-

reader interaction (e.g. Vande Kopple 1985, Crismore 1989, Hyland 2005), or a narrow approach focusing on the aspects of text organization and elements referring to the text itself (e.g. Mauranen 1993, Ädel 2006)[1]. In his metadiscourse model, Hyland (2005: 52) characterizes code glosses 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 (2007) also distinguished two broad sub-functions of code glosses, 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" (ibid.: 269). In academic writing such connections are often signalled by what Hyland calls 'reformulation markers' (*that is*, *or*, *put another way*) or simply by brackets. An example from the SciELF corpus is as follows:

(1) *To accomplish a particular aesthetic impact, the writer exploits distinctive communicative strategies. <u>In other words,</u> the author uses various linguistic means to create the effect connected to the central meaning of the text, or to the meaning, on which the author focuses on various levels to variable extent throughout the whole text.* (SSH16)

It should be noted that Hyland's conception of reformulation is broad. It covers not only the traditional reformulation markers, which are grammaticalized (such as *i.e.*, *that is*, *namely*, *in other words*; see e.g. Cuenca 2003), but also various lexical means used to rephrase or explain what has been said (for example, *this means*, *this can be defined as*, etc.).

The second broad function of code glosses is exemplification. According to Hyland (2007: 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 is a common way of making arguments clear and persuasive. What follows is an example from the SciELF:

(2) *Control can be defined in either financial or operational terms. <u>For example,</u> an equity share of 60% is considered to be financial control, and thus the parent company will consolidate 100% of the GHG emissions of the subsidiary.* (SSH50)

A number of code glosses can be classified as 'appositive conjuncts' in Quirk et al.'s (1985) terms, forming a scale. Appositives such as *that is*, *namely* or *in other words* express the semantic relationship of equivalence, and are most appositive on the scale. The indicators of exemplification, in which the second appositive exemplifies the reference of the more general term in the first expression, are the least appositive on the scale (ibid: 1308). Similarly, these expressions are treated as 'adverbials of apposition' in Biber et al. (1999: 876), where "the second unit of text is to be treated either as equivalent to or included in the preceding unit".

However, it seems that not all appositive conjuncts can be considered reflexive, i.e. not adding anything to the propositional content.[2] Hyland's (2007) list of potential code glosses expressing reformulation includes adverbials *especially*, *particularly* and *in particular*. Nevertheless, the status of these restrictive adverbials seems controversial since their primary function is not to reformulate. They emphasize the importance of one part of the proposition "by restricting the truth value of the proposition either primarily or exclusively to that part" (Biber et al. 1999: 556). These expressions are not considered reflexive in this study since they alter the truth value of the proposition and are therefore treated as propositional. It is interesting to note that even Hyland did not include them in the list of code glosses in some other studies (e.g. Hyland 2005, Hyland 2012).

Exemplifying expressions are included in low-explicit reflexivity in Mauranen's model (1993: 181) since they are considered both conceptual categories and linguistic ones. Low-explicit reflexivity includes expressions which "organise the text and indicate functions of its parts without explicitly referring to the communication process" (ibid.: 180). Therefore, the indicators of exemplification are considered reflexive in this study as they explicitly introduce an example (they can be also treated as 'discourse labels', see Ädel 2006).

3   Data and methodology

The research is based on the comparison of two corpora. The first one is the SciELF corpus, which consists of 150 research articles written by L2 users of English, totalling 759,300 words. The articles have not undergone professional proofreading or checking by English native speakers, and most of them are final drafts of unpublished manuscripts. As for the disciplines, the SciELF corpus is divided into sciences (Sci=78 articles) and social sciences and humanities (SSH=72 articles). Since disciplinary variation in the use of metadiscourse is high (e.g. Hyland 2005), I decided to compare only social sciences and humanities articles.

The authors of the articles come from ten different L1 backgrounds – Finnish, Czech, French, Chinese, Spanish, Russian, Swedish, Italian, Portuguese and

Romanian. The number of articles written by each language group vary, but the aim of ELF research is not to do L1-based comparisons but to investigate how people from different language backgrounds use English as a lingua franca in academic settings. The total number of words in the corpus is 759,300. The Sci part contains 326,463 words and the articles are mostly drawn from natural sciences (79%) and medicine (18%). The SSH part contains 432,837 words and the articles are drawn from social sciences (45%), humanities (34%), and behavioural sciences (21%).

As mentioned above, the analysis has been carried out on the SSH part of the corpus including social sciences and humanities (SSH=72 articles). The disciplines represented in the corpus are social policy, educational sciences, psychology, economics, classical philology, linguistics, philosophy, law, theatre studies, anthropology, information sciences, sociology, history, urban design, art history and literary studies.

In order to investigate the use of code glosses in written academic ELF, a comparable corpus was compiled. The second corpus is composed of 42 English-medium articles, which were published in prestigious academic journals. The papers were written by Anglophone authors based at British, American or Australian universities. Although it cannot be guaranteed that all of the authors are native speakers of English, it is possible to assume that their L1 is English based on their names, affiliations and sometimes CVs. The journals were selected according to their impact factor (Web of Science data, 2018) or SCImago rankings (Scopus, 2018) and the categories were chosen in order to mirror the disciplines of the SciELF corpus (or specifically its social sciences and humanities component). For example, the SciELF corpus contains articles from educational sciences, so the category Education and Educational Research in the Web of Science database was searched and two journals with a high impact factor were chosen (and also based on the consultation with colleagues from the field) – *The Internet and Higher Education* (Impact Factor 5.284, 2018) and *Educational Researcher* (Impact factor 3.386, 2018). So far 42 articles have been selected for inclusion in the comparable corpus, totalling 389,558 words, with 9,275 average word count per article. The articles are on average longer than the papers in the SciELF corpus, which was mainly caused by the length of economic papers published in high-impact journals; the average word count in economic papers is 13,112. All the papers were published between 2016-2018 and they are single-authored. Table 1 shows the descriptive figures for both corpora:

	SciELF-SSH corpus	ENL-SSH corpus
Number of articles	72	42
Number of words	432,837	389,558
Average words/article	6,012	9,275
Texts	final drafts of unpublished RAs	published RAs
Disciplines	educational sciences, psychology, economics, classical philology, linguistics, philosophy, law, theatre studies, anthropology, sociology, history, urban design, literary studies, information sciences, art history, social policy	educational sciences, psychology, economics, classical philology, linguistics, philosophy, law, theatre studies, anthropology, sociology, history, urban design, literary studies

Table 1: Composition of the SciELF corpus and the ENL corpus

In preparing the texts for inclusion in the corpus the SciELF Corpus Manual was followed so that the two corpora would be comparable. That is why bibliographic references as well as long stretches of foreign text and block quotes were omitted in the plain text files (which are used for concordancers), while abstracts of the articles were kept.

The texts were 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 2018), using the concordancer AntConc. Unlike Hyland (2007) I focused only on explicit textual devices, so I did not consider punctuation markers, such as parentheses, as metadiscourse markers. Then, all cases were examined in context to ensure 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. Furthermore, code glosses which appeared in quotations and examples were omitted since they were not regarded as the manifestation of the authorial voice and thus part of metadiscourse. After the quantitative analysis, the functions of code glosses were considered and the differences between the two corpora discussed.

4 Results and discussion

The two corpora were at first contrasted in terms of the frequency of occurrence of code glosses. The results of the quantitative analysis are summarized in Table 2.

	SciELF-SSH corpus			ENL-SSH corpus		
	Total no.	Freq. per 10,000 words	% of total code glosses	Total no.	Freq. per 10,000 words	% of total code glosses
Markers of reformulation	650	15.0	42.8	484	12.4	34.1
Markers of exemplification	870	20.1	57.2	934	24.0	65.9
Code glosses (TOTAL)	1,520	35.1	100.0	1,418	36.4	100.0

Table 2: Distribution of code glosses in the SciELF-SSH corpus and the ENL-SSH corpus

The overall frequency of code glosses in the two corpora was similar – 35.1 items per 10,000 words in the ELF corpus and 36.4 per 10,000 words in the ENL corpus. The difference is not statistically significant $(G^2=0.94)^3$. Furthermore, in both corpora exemplification predominates over reformulation. This corroborates Hyland's findings (2007: 272) that exemplification plays a key role in soft knowledge fields, representing "a heavier rhetorical investment in contextualisation" since researchers in the soft fields cannot assume that the background to research or criteria for establishing results are agreed on by all readers. Instead, they need to continuously re-establish the context, and exemplification is a convenient way of doing so. The distribution of functions was somewhat different, since in the ELF corpus the ratio of exemplifying markers and reformulation markers was 57.2 per cent to 42.8 per cent, while in the ENL corpus the ratio was 65.9 per cent to 34.1 per cent. While ELF writers restated or elaborated their utterances using different words more often than ENL writers, ENL writers used more examples. The differences are statistically significant $(G^2=10.04; p<0.01$ for reformulation and $G^2=14.02; p<0.001$ for exemplification); however, the corpus is too small to draw any general conclusions.

4.1 Reformulation in research articles

The first part of the analysis focused on reformulation markers. Table 3 shows the distribution of specific code glosses expressing reformulation in the SciELF and the ENL corpora.

REFORMULATION				
Marker	**SciELF-SSH corpus**		**ENL-SSH corpus**	
	Total/Frequency per 100,000 words	% of reformulation markers	Total/Per 100,000 words	% of reformulation markers
i.e.	177/40.9	27.2	83/21.3	17.1
that is	81/18.7	12.5	60/15.4	12.4
mean*	71/16.4	10.9	69/17.7	14.3
in other words	56/12.9	8.6	36/9.2	7.4
namely	49/11.3	7.5	13/3.3	2.7
called	46/10.6	7.1	17/4.4	3.5
define*	39/9.0	6.0	48/12.3	9.9
specifically	35/8.1	5.4	55/14.1	11.4
refer* to	31/7.2	4.8	35/9.0	7.2
or x	30/6.9	4.6	48/12.3	9.9
known as	22/5.1	3.4	9/2.3	1.9
put it	6/1.4	0.9	6/1.5	1.2
more precisely	5/1.2	0.8	0/0	0
termed	2/0.5	0.3	5/1.3	1.0
Total	650/150.2	100.0	484/124.2	100.0

Table 3: Frequency of specific code glosses expressing reformulation in the SciELF-SSH and ENL-SSH corpora

The most common device functioning as a reformulation marker in the SciELF-SSH corpus is the abbreviation i.e., followed by *that is*, *mean** (*mean*, *means*, *meant*, *meaning*), in *other words* and *namely*. In the reference corpus the order is similar, with the abbreviation i.e. followed by *mean**, *that is*, *specifically*, *define** (*define*, *defines*, *defined*, *defining*) and *or x*. What proved interesting is that although the abbreviation *i.e.* is the most frequent reformulation marker in both corpora, it is much more frequent among ELF speakers (40.9 to 21.3 items per 100,000 words). While in the ENL corpus it accounts for 17.1 per cent of all reformulation markers, in the SciELF corpus it is 27.2 per cent. A log-likelihood test on the number of i.e. confirmed that there is a statistically significant difference between the SciELF and the ENL corpora (G^2=25.58; p < 0.0001).

As far as punctuation is concerned, 77 out of the total 83 instances of the abbreviation *i.e.* (93%) occurred in parentheses in the ENL-SSH corpus signalling that it introduced background information rather than the main idea. This tendency is likely to be influenced by the APA style guide which states that standard Latin abbreviations should only be used in parenthetical material (and the English translations in non-parenthetical material) (2010: 108). In the

SciELF corpus only 63 out of 177 instances of i.e. occurred in parentheses (36%), pointing to a greater variation in use.

Another difference between the two corpora can be identified in the use of *namely*, which occurred significantly more often in the SciELF-SSH corpus (G^2=18.65; $p < 0.0001$). This marker was used to specify a previous element, and it often introduced a list, especially in the SciELF corpus:

(3)   *As we saw earlier, there are three types of bilingual instruction, <u>namely</u> immersion bilingual instruction, maintenance bilingual instruction and transitional bilingual instruction.* (SSH07)

A higher density of this marker in the SciELF corpus might suggest that ELF users make such specifications more explicit in order to get the meaning across and thus try to be more cooperative.

As for the reformulation marker *that is*, it is interesting to note that its longer version *that is to say* occurred only in the ELF corpus (27 occurrences, 6.2 per 100,000 words). The phrase was not attested in the ENL corpus at all. Incidentally, the frequency in the Academic sub-corpus of the British National Corpus is 1.8 per 100,000 words, which is still significantly lower than in the SciELF. One explanation for this is provided by Murillo (2018: 245) who argues that the higher incidence in the ELF corpus may be accounted for by the existence of similar markers in Romance languages (e.g. French *c'est-a-dire* and Spanish *es decir*).

Hyland (2007) distinguishes a range of functions that reformulation markers fulfil. They can either expand the original, by explanation or implication, or reduce it by paraphrase or specification. Reformulations used for the expansion restate an idea in order to widen the sense in which the writer intends it to be understood. This is achieved by offering an explanation (including clarifications for a technical term) or by drawing an implication of a prior segment (e.g. using *in other words*). Reformulations reducing the original idea serve to narrow the scope of interpretation by either paraphrase ("restating an idea in different words to provide a summary") or specification (introduced by *specifically*, *more accurately* etc.) (Hyland 2007: 276).

It should be noted that Hyland's (2007: 277) functions represent general meaning relations among the units, i.e. they are semantic. The most common function of reformulation in his study proved to be specification. This function occurred frequently in SciELF corpus as well, for example in RA introductions where research aims were specified (4).

(4) *This essay explores some paintings of the Fernando Botero's "Abu Ghraib" series (2003) in light of theoretical approaches from a postcolonial perspective, <u>more specifically</u> in matters concerning cultural identity explained in terms of hybridity and orientalism.* (SSH66)

Another function of reformulation which occurred repeatedly is defining concepts (signalled lexically by *define*, *refer* to or by fixed expressions *i.e.*, *that is*). Being aware of the fact that the terminology is not completely unified, especially within soft disciplines, the writers defined concepts in order to show in which sense they are used for the purposes of their study (5). The definitions often include a citation or a reference to a source.

(5) *For the purpose of this study the term inference <u>refers to</u> "any information about events, relations, and so on that the reader adds to the information that is explicitly presented in the text" (van den Broek, et al., 1995, p.353).* (SSH48)

Conversely, reformulation markers also provide a term or a designation for a concept already explained (signalled by *called*, *known as*, *referred to as*, *termed*, *i.e.*) (6).

(6) *The two areas were connected by an all-encompassing roof, a platform one kilometre in length and 150 metres wide (<u>called</u> the 'Symbol Zone').* (SSH64)

In the following example, *that is* introduces an explanation of a previous author's statement (7). Explanations sometimes followed citations, as in Example (8).

(7) *In addition to rupturing the logical hypotactic construction, by translating the hypotaxis into an adjunct,* jovialmente *forms a misleading picture in terms of the meaning provided by* laughing. *<u>That is</u>, it only expresses a person who is happy and friendly, and not a person who feels an exquisite pleasure for having given such a brilliant and dissenting opinion on moral edicts.* (SSH47)

(8) *According to this treaty, "Russian soldiers were to be stationed in Georgia, and the king was to conduct his foreign policy in conformity with Russian preferences" (Suny 58). <u>In other words</u>, the Georgian kingdom was to be internally independent, while its foreign policy was to be in the hands of the Russian tsar.* (SSH17)

Reformulation markers were also used to draw conclusions. In Example (9) the writer makes a conclusion from a previous paragraph, pointing towards a gap in research and thus providing a rationale for his/her study. Writers thus make sure that readers will interpret a given message correctly.

(9)   *To my knowledge, verbal imageability in narrative, i.e. readers' vicarious hearing
        of the words on a page, has not been investigated empirically. A few empirical
        studies have been conducted focusing on the effects of specific phonemic patterns
        on interpretation and affective response (for a review, see Miall 2006, 173-188).
        However, these studies are inconclusive as to whether these patterns are assumed
        to inform the reader's consciousness directly.*
        *In other words, extant empirical findings regarding imageability in narrative are
        neither surprising nor many.* (SSH11)

However, reformulation markers need not only explain, define or clarify, but
they can also support writer's arguments. In the following example, the writer
provides specific data to give evidence of his argument (10).

(10)   *The motivation and growth of online education varies according to institution
        type, but a large percentage of the proliferation of online offerings can be
        explained by institutions moving from offering the occasional online course to
        providing complete fully online degree programs to their students. Specifically, the
        percentage of American higher education institutions offering fully online degree
        programs increased from 34.5% in 2002 to 62.4% in 2012 (Allen & Seaman,
        2014).* (ASSH15)

Therefore, the use of reformulation in both corpora shows that unlike in speech
where reformulation markers often function as repairs (or self-rephrasing which
is especially important among ELF speakers (Mauranen 2012)), in a written,
planned discourse they have to be considered as deliberate communicative
strategies used by the writer. One of these strategies is to support writer's
arguments and to project a credible writer persona.

4.2 Exemplification in research articles

The second broad function of code glosses is exemplification. Exemplification
is an important part of argumentation in academic texts as writers can support
their arguments by providing specific examples to general statements. It is also
a process through which the meaning is clarified or explained, so it reflects the
writer's effort to communicate the message in such a way as to make it coherent and
understandable for the reader. The writer aims to meet the reader's expectations
and processing needs, and using code glosses is one way of achieving it. As
Mauranen put it, reflexive language is "highly interactive, whether couched in
monologic or dialogic form" (2010: 16).

The distribution of specific exemplification markers in both corpora is given
in Table 4. Four expressions were chosen for the analysis (*for example*, *e.g.*, *for*

instance, *such as*) since they were found to be the most frequent exemplification markers in social sciences by previous studies (e.g. Hyland 2007: 278).

EXEMPLIFICATION				
	SciELF-SSH corpus		**ENL-SSH corpus**	
Marker	Total/Frequency per 100,000 words	% of exemplification markers	Total/Frequency per 100,000 words	% of exemplification markers
such as	357/82.5	41.0	341/87.5	36.5
e.g.	237/54.8	27.2	256/65.7	27.4
for example	185/42.7	21.3	247/63.4	26.5
for instance	91/21.0	10.5	90/23.1	9.6
Total	870/201	100.0	934/239.7	100.0

Table 4: Frequency of specific code glosses expressing exemplification in the SciELF-SSH and ENL-SSH corpora

The total number of exemplification markers was higher in the ENL-SSH corpus than in the SciELF-SSH and the difference is statistically significant (G^2=14.02; p < 0.001). The main difference concerned the marker *for example*, whose frequency was 63.4 per 100,000 words in the ENL corpus and 42.7 per 100,000 words in the SciELF. However, the proportional distribution of individual markers in both corpora is similar, with *such as* being the most frequent marker. This is consistent with Hyland (2007), who found out that *such as* was favoured in most disciplines.

Similarly to the abbreviation *i.e.*, there was a strong tendency for the marker *e.g.* to occur in parentheses. In fact, all but three instances of *e.g.* appeared in parentheses in the ENL corpus (253 cases), showing that examples signalled by *e.g.* often illustrate background information. The number is slightly lower in the SciELF (173 out of 237 instances, which makes 73% of the total), pointing towards a greater variation, but this tendency is still prominent. Moreover, within the parentheses the marker *e.g.* often introduced a reference to a different source by means of which writers supported their arguments (79 cases in the SciELF corpus and 138 cases in the ENL corpus, i.e. about a half of all the examples in the parentheses). Writers either give examples of studies which state the facts upon which the writers build their arguments (11) or they give examples of studies which dealt with the problem discussed (12). Writers can also directly refer their readers to further sources dealing with the topic (introduced by *see e.g.*), thus contextualizing their research within a discipline as a whole (Dontcheva-Navratilova 2016: 57).

(11)   *It is a well-established fact that beyond the age of puberty, it is virtually impossible to acquire a standard pronunciation in the target language (<u>e.g.</u> Lenneberg: 1967; Scovel: 1994).* (SSH09)

(12)   *More recently, however, there has been a significant increase of attention to the ways in which language is enmeshed with other semiotic resources in constructing meanings in communication (<u>e.g.,</u> Goodwin 2000; Lemke 2002).* (ASSH01)

Considering the exemplification marker *for instance*, the results show that it was not distributed evenly across the texts. The SciELF-SSH corpus included 91 instances in 25 texts (out of the total 72). However, 51 cases of *for instance* occurred only in three texts, which suggests that the use of this marker is a matter of author's personal style (see Biber et al. 1999). The three authors did not use the marker *for example* to introduce their examples (with the exception of four cases in brackets), they preferred exclusively *for instance*. Similarly, the ENL-SSH corpus included 90 cases of *for instance* in 19 texts (out of 42), but the overall distribution in individual texts was quite even.

The functions of examples in both corpora ranged from explaining or clarifying writers' statements (e.g. in methodology) to illustrating their points and thus supporting their arguments. This was often achieved by means of citations since supporting the statements by references to credible sources is an important persuasive strategy, helping to situate the research in a larger context of a discipline (13).

(13)   *From the standpoint of an epistemological debate, knowledge and beliefs may be seen as separate entities in that knowledge is closer to truth than beliefs are. <u>For example</u>, Fenstermacher (1994) distinguished beliefs from knowledge by emphasising beliefs as personal values which may not have epistemic merit and knowledge as factual proposition.* (SSH06)

4.3 Non-canonical uses of code glosses by ELF writers

As far as the realizations of code glosses in the SciELF corpus are concerned, some non-standard formulations occasionally appeared. For example, two authors used the structure called as (14), which might be called approximation. Example (15) includes an infinitive instead of a subordinate clause which would make the sentence more comprehensible (*It means that it is appropriate for…*).

(14)   *In the field of Terminology and Computer science, for example, researchers have tried to automatise the extraction of knowledge-rich contexts and terminological information from texts by using terminological markers, also <u>called as</u> knowledge probes […].* (SSH22)

(15)   *Susan Sontag writes that since the camera was invented (1839) the photograph socialises with death. Because of its speed, an image captured in a second, a photograph exceeds painting for its "documentary" force. <u>It means to be appropriate of a "capturing moment of death".</u>* (SSH66)

However, these examples were isolated as standard forms clearly prevailed. The only deviation, or non-canonical use of code glosses, which occurred several times across texts was the use of a restrictive adverb *especially* at the beginning of a sentence, followed by a subject and a predicate (16). Although *especially* is not regarded as a reformulation marker in this study, Hyland includes it in his broad conceptualisation of reformulation. Therefore, the corpus was briefly searched for these markers as well, and it yielded this pattern. The online *Longman Dictionary of Contemporary English* states that "*Especially* is not used at the start of a sentence before the subject", such as **Especially young people are being affected.*[4] This structure was used seven times in the whole SciELF corpus (both in natural and social sciences) by five different authors, and it twice occurred at the beginning of a clause separated by a comma in the function of a sentence adverb (17).

(16)   <u>*Especially*</u> *Törnqvist has emphasised the intertextual, especially the Biblical, references of The Burned Site.* (SSH70)

(17)   *Different factors contributed to that substantial growth. <u>Especially</u>, several federal acts were passed and a safe regulatory environment was created for development of mortgage lending.* (Sci65)

This structure was not attested in the ENL-SSH corpus. When I consulted the Academic sub-corpus of the British National Corpus, I found only one hit of *Especially* followed by a comma, which on closer inspection appeared in a research article authored by French researchers.[5] With the exception of this case the BNC did not attest any sentences beginning with *Especially* followed by a subject and a predicate.

5   Conclusion

The analysis of code glosses in the SciELF corpus proved that this type of discourse reflexivity is an indispensable part of ELF written discourse. Their frequency in the SciELF-SSH corpus was almost the same as in the comparable corpus of published research articles written by Anglophone writers. The results have also shown that ELF is in many respects similar to ENL as writers

must recognize rhetorical expectations and preferences of their readers to be successful in their argumentation, and one way of achieving this is to provide a reformulation or elaboration of what has been said.

One of the tendencies observed in the ELF corpus was simplification. Previous research has identified lexical simplification in academic ELF speech, manifested for example in the tendency to higher representation of top-frequency items in ELFA (Mauranen 2017: 240). This process was also shown in reformulation markers, with the most frequent marker in the ENL corpus *i.e.* being even more prominent among ELF users (accounting for 27% of all reformulation markers). It suggests that some of the most frequent expressions in ENL are even more frequent in ELF.

In some respects, the ENL corpus was more homogenous since it represents research articles already published in prestigious journals. For example, the parenthetical use of abbreviations *i.e.* and *e.g.* is partly given by the house rules of the journals using APA or a similar norm. Still, the more prominent variation in their use in the SciELF can point towards complexifying tendencies among ELF users. Such tendencies can also be seen in the use of some non-standard features (the use of prepositions – *called as*, word order – *especially*). This increased variability is making the system more complex in the sense that "the new elements entering it […] have the potential of throwing the system (or any subsystem in it) off balance" (Mauranen 2017: 244). This study of written academic ELF thus corroborates the assumption that ELF manifests tendencies of simplification as well as complexification resulting from the higher-order form of language contact (ibid.: 246).

As for discourse reflexivity as a manifestation of enhanced explicitness, the ELF and the ENL corpora were quite similar in the frequency of code glosses. However, the SciELF-SSH corpus attested higher number of reformulation markers, which are considered more explicitly reflexive than exemplification markers. Future research is needed, especially on highly explicit reflexivity, to ascertain general tendencies in written academic ELF. Moreover, unlike in spoken discourse, code glosses in written texts have to be regarded as deliberate communicative strategies, bringing an interactive element to discourse and supporting writer's arguments.

Acknowledgement

Notes

1. The two approaches were first distinguished by Mauranen (1993). For a detailed discussion, see, for example, Ädel (2006).
2. According to Hyland (2005: 38), one of the key principles of metadiscourse is that it is distinct from propositional aspects of discourse. Even though this distinction has been challenged (see e.g. Guziurová 2018), I follow it here for methodological reasons.
3. The log-likelihood test was carried out with an online calculator created by Paul Rayson (http://ucrel.lancs.ac.uk/llwizard.html, accessed 30 June 2019). Following Carey (2013), each occurrence of a code gloss was treated as a single observation.
4. https://www.ldoceonline.com/dictionary/especially, accessed 30 June 2019.
5. Lehy, T., Cadiot, G., Mignon, M., Ruszniewski, P. and Bonfils, S. (1992) 'Influence of multiple endocrine neoplasia type 1 on gastric endocrine cells in patients with the Zollinger-Ellison syndrome.' *Gut 33*(9), 1275-1279.

References

Ädel, A. (2006) *Metadiscourse in L1 and L2 English.* Amsterdam and Philadelphia: John Benjamins.

American Psychological Association. (2010) *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Anthony, L. (2019) AntConc (Version 3.5.8) [Computer Software]. Tokyo: Waseda University. Available from 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. and Borelli, C. (2018) 'Publishing in English: ELF writers, textual voices and metadiscourse.' In: Mur-Dueñas, P. and Šinkūnienė, J. (eds) *Intercultural Perspectives on Research Writing.* Amsterdam and Philadelphia: John Benjamins. 217-235.

Carey, R. (2013) 'On the other side: Formulaic organizing chunks in spoken and written academic ELF.' *Journal of English as a Lingua Franca 2*(2), 207-228.

Connor, U. (1996) *Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing.* Cambridge: Cambridge University Press.

Crismore, A. (1989) *Talking with Readers: Metadiscourse as Rhetorical Act.* New York: Peter Lang.

Cuenca, M. J. (2003) 'Two ways to reformulate: A contrastive analysis of reformulation markers.' *Journal of Pragmatics 35*(7), 1069-1093.

Dontcheva-Navratilova. O. (2016) 'Rhetorical functions of citations in linguistics research articles: A contrastive (English-Czech) study.' *Discourse and Interaction 9*(2), 51-74.

Guziurová, T. (2018) *Metadiscourse in Undergraduate Textbooks and Research Articles in Linguistics.* Ostrava: Pedagogická fakulta Ostravské univerzity.

Hyland, K. (2005) *Metadiscourse.* London and New York: Continuum.

Hyland, K. (2007) 'Applying a gloss: Exemplifying and reformulating in academic discourse.' *Applied Linguistics 28*(2), 266-285.

Hyland, K. (2012) *Disciplinary Identities: Individuality and Community in Academic Writing.* Cambridge: Cambridge Applied Linguistics.

Hyland, K. (2017) 'Metadiscourse: What is it and where is it going?' *Journal of Pragmatics 113*, 16-29.

Hyland, K. and Milton, J. (1997) 'Qualification and certainty in L1 and L2 Students' Writing.' *Journal of Second Language Writing 6*(2), 183-205.

Mauranen, A. (1993) *Cultural Differences in Academic Rhetoric: A Textlinguistic Study.* Frankfurt am Main: Peter Lang.

Mauranen, A. (2007) 'Discourse reflexivity and international speakers: How is it used in English as a lingua franca?' *Jezik in slovstvo 52*(3-4), 1-19.

Mauranen, A. (2010) 'Discourse reflexivity – A discourse universal? The Case of ELF.' *Nordic Journal of English Studies 9*(2), 13-40.

Mauranen, A. (2012) *Exploring ELF. Academic English Shaped by Non-native Speakers.* Cambridge: Cambridge University Press.

Mauranen, A. (2017) 'A glimpse of ELF.' In: Filppula, M., Klemola, J., Mauranen, A. and Vetchinnikova, S. (eds) *Changing English: Global and Local Perspectives.* Berlin, Boston: De Gruyter. 223-254.

Mauranen, A., Hynninen, N. and Ranta, E. (2016) 'English as the academic lingua franca.' In: Hyland, K. and Shaw, P. (eds) *The Routledge Handbook of English for Academic Purposes.* Abingdon: Routledge. 44-55.

Mur Dueñas, P. (2018) 'Exploring ELF manuscripts: An analysis of the anticipatory it pattern with an interpersonal function.' In: Mur-Dueñas, P. and Šinkūnienė, J. (eds) *Intercultural Perspectives on Research Writing.* Amsterdam and Philadelphia: John Benjamins. 277-297.

Murillo, S. (2018) 'Not the same, but how different? Comparing the use of reformulation markers in ELF and in ENL research articles.' In: Mur-Dueñas, P. and Šinkūnienė, J. (eds) *Intercultural Perspectives on Research Writing.* Amsterdam and Philadelphia: John Benjamins. 237-253.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language.* London: Longman.

Rowley-Jolivet, E. (2017) 'English as a Lingua Franca in research articles: The SciELF corpus' *ASp 71*, 145-158.

SciELF 2015. *The SciELF Corpus.* Director: Anna Mauranen. Compilation manager: Ray Carey.

Sinclair, J. (2005) 'Language as a string of beads: Discourse and the M-word.' In: Tognini-Bonelli, E. and Del Lungo Camiciotti, G. (eds) *Strategies in Academic Discourse.* Amsterdam and Philadelphia: John Benjamins. 163-168.

Valero-Garcés, C. (1996) 'Contrastive ESP rhetoric: Metatext in Spanish-English economics texts.' *English for Specific Purposes 15*(4), 279-294.

Vande Kopple, W. (1985) 'Some exploratory discourse on metadiscourse.' *College Composition and Communication 36*(1), 82-93.

Vassileva, I. (2001) 'Commitment and detachment in English and Bulgarian academic writing.' *English for Specific Purposes 20*, 83-102.

Vetchinnikova, S. (2014) *Second Language Lexis and the Idiom Principle.* Helsinki: University of Helsinki.

Wu, X., Mauranen, A. and Lei, L. (2020) 'Syntactic complexity in English as a lingua franca academic writing.' *Journal of English for Academic Purposes 43*, 100798.

Tereza Guziurová is Assistant Professor of English Linguistics at the Faculty of Education, University of Ostrava, Czech Republic. She specializes in discourse analysis and genre analysis. Her research interests include metadiscourse in academic genres and the aspects of writer-reader interaction in academic texts

in general. Recently her research has focused on English as a Lingua Franca in academic settings. In 2018 she published a monograph *Metadiscourse in Undergraduate Textbooks and Research Articles in Linguistics*.

Address: Tereza Guziurová, Department of English Language with Didactics, Faculty of Education, University of Ostrava, Fráni Šrámka 3, 709 00 Ostrava, Czech Republic. [e-mail: tereza.guziurova@osu.cz]