

Beyond words: Exploring the use of code glosses in L2 English academic discourse by Serbian and Croatian novice writers

Jagoda Topalov, Vesna Bogdanović & Biljana Radić-Bojanić

University of Novi Sad (Serbia)

jagoda.topalov@ff.uns.ac.rs, vesna241@uns.ac.rs,

radic.bojanic@ff.uns.ac.rs

Abstract

Code glosses (e.g., *in other words*, *namely*, *for instance*, *such as*) are metadiscourse markers that restate propositional content while reflecting the writer's attitude towards the subject matter and the audience, thereby projecting the writer's relationship with both. Code glosses can be (mis)used to influence the audience and their opinion. Nevertheless, their appropriate use is a skilled knowledge that has to be acquired. The paper examines whether graduate students in English language and literature from two universities whose L1 are Serbian and Croatian can properly recognize and utilize code glosses in order to create comprehensible and reader-friendly theses. Following Hyland's (2005, 2007) framework, this study employed a quantitative analysis to identify and examine code glosses in a corpus comprising the Discussion sections of MA theses written by students. The analysis was conducted in contrast to a reference corpus of Discussion sections from MA theses by students from American and British universities. In this study, the results of statistical tests reveal that novice L2 writers demonstrate a propensity for reformulation, while underusing exemplification and mostly rely only on a small number of familiar markers. This may be a result of insufficient training in academic writing, differences across disciplines, variations in culture and language, and a lack of understanding regarding the roles and significance of code glosses in academic discourse.

Keywords: code glosses, metadiscourse, L2 novice writers, quantitative discourse analysis, marker underuse, marker overuse.

Resumen

Más allá de las palabras: explorando el uso de glosas de código en el discurso académico en inglés como L2 de escritores noveles serbios y croatas

Las glosas de código (p. ej., *en otras palabras, a saber, por ejemplo, como*) son marcadores metadiscursivos que reformulan el contenido proposicional y reflejan la actitud del autor hacia la materia de estudio y hacia la audiencia, proyectando así la relación del emisor con ambas. Estas glosas pueden emplearse —a veces indebidamente— para influir en la audiencia y en sus opiniones. Sin embargo, su uso apropiado constituye una destreza que debe adquirirse. Este artículo examina si los estudiantes de posgrado de lengua y literatura inglesas de dos universidades, cuya lengua materna es el serbio y el croata, son capaces de reconocer y utilizar adecuadamente las glosas de código para elaborar trabajos fin de máster comprensibles y orientados al lector. Siguiendo el marco de Hyland (2005, 2007), este estudio emplea un análisis cuantitativo para identificar y examinar las glosas de código en un corpus integrado por las secciones de discusión de los trabajos fin de máster escritos por estos estudiantes. Estos textos se comparan con un corpus de referencia compuesto por secciones de discusión de trabajos fin de máster elaborados por estudiantes de universidades estadounidenses y británicas. En este estudio, los resultados de las pruebas estadísticas revelan que los escritores noveles de L2 tienden a la reformulación, mientras que infrautilizan la ejemplificación y dependen principalmente de un número reducido de marcadores que les resultan familiares. Esto puede deberse a una formación insuficiente en escritura académica, a diferencias entre disciplinas, a variaciones culturales y lingüísticas y a una comprensión limitada de las funciones y la relevancia de las glosas de código en el discurso académico.

Palabras clave: glosas de código, metadiscurso, escritores noveles de L2, análisis cuantitativo del discurso, infrautilización de marcadores, sobreutilización de marcadores.

1. Introduction

Academic writing has to be factual, convincing and easy to read and interpret (Biber & Gray, 2010; Hyland & Jiang, 2017; Swales & Feak, 2004). Hence, propositional material is of essence for presenting the author's research and disseminating knowledge to a wider academic public. However, apart from presenting their ideas, authors also have to accommodate their readers. Therefore, rhetoric choices made by the authors to negotiate appropriate engagement and explicitness (Hyland, 2007) are essential for communication

with the readers. When writing academic papers, authors have to study the audience in their discourse community and consider their knowledge base in order to produce effective writing (Alghazo et al., 2021). After potentially predicting the readers' subject knowledge, academic writers have to be prepared to expand, rework and/or exemplify their ideas to produce coherent and reader-friendly texts. Metadiscourse, therefore, has a key role in academic writing as language that helps organize discourse, engage readers and express the writer's stance.

In the literature on metadiscourse, the rhetoric devices for providing examples and restating information are known as code glosses (Hyland, 2005). They are used for reformulation and exemplification (Hyland, 2007) as types of explanation (Calsamiglia & Van Dijk, 2004), i.e. a rhetoric structure to adjust information to the readers' knowledge and information needed. They are also regarded as a strategy of recontextualization (Luzón, 2013). Additionally, writers use code glosses to support their position and increase their communicative efficiency (Kafes, 2022). Thus, writers express their judgement on the readers, they state their stance towards the discourse community, and they establish the relationship between the message and the readers (Hyland, 2005).

In comparison to other metadiscourse devices, code glosses are found to be least frequently used interactive metadiscourse markers (Hyland & Jiang, 2018). They are present in research articles, but are also present in undergraduate textbooks (Hyland, 2005; Guziurová, 2022) in order to clarify unfamiliar concepts or terminology. Since students need more assistance in comprehension, code glosses are used to provide definitions and additional information to clarify issues in textbooks (Liu et al., 2023). In order for innovative and educational content to reach a wider audience, code glosses are typical for research discourses and science journalism (Hyland, 2009). As a consequence, these discursual practices have the option of avoiding jargon, or offering jargon and terminology with code glosses as a communicative strategy. In addition, code glosses can be found, for example, in academic blogs (Luzón, 2013; Bondi, 2020), three minute thesis presentations (Liu et al., 2013), argumentative essays (Aull & Lancaster, 2014), research articles (Dehghan & Chalak, 2016; Kafes, 2022), ESP textbooks (Walková, 2020), oral presentations (Nausa Triana, 2019; Siepmann, 2023), etc. Additionally, they are present in both hard and soft disciplines (Hyland, 2005; 2007) from papers on politics to papers on applied linguistics (Safari, 2018), where both mentioned disciplines rely more on reformulations than exemplifications.

While most studies examine code glosses as part of broader metadiscourse frameworks, research that isolates and analyzes them independently remains relatively limited. However, this does not imply that such studies are less representative or methodologically unsound. On the contrary, a focused approach allows for a more precise investigation of how code glosses function in academic discourse, offering insights that may be overlooked in studies that analyze metadiscourse as a whole. By examining code glosses separately, this study contributes to a more nuanced understanding of their role in L2 writing, particularly in underexplored linguistic and cultural contexts. In line with this, the use of code glosses is compared between L1 English academic writing and Iranian L2 English (Dehghan & Chalak, 2016), Spanish and Catalan writing (Cuenca & Bach, 2007), Albanian and German (Edusei, 2015), Czech (Guziurová, 2020; 2022) and Turkish (Kafes, 2022). The use of code glosses is also investigated in comparison between expert and novice writers (Guziurová, 2022; Kafes, 2022). For example, novice academic writers may use code glosses to demonstrate their persuasion skills in presenting their claims to the readers and their discourse community, facilitating the understanding of the message (Edusei, 2015). Moreover, oral presentations by graduate students (Nasau Triana, 2019; Liu et al., 2023) demonstrate that students contextualize the content they have to present by utilizing code glosses more than other interactive metadiscourse devices.

Following this line of research, the aim of the paper is to investigate the use of code glosses by graduate students in English linguistics whose L1 are Serbian and Croatian, two typologically very similar languages, contrasting it with the use of code glosses by native English student writers from British and American universities. Considering the relative inexperience of L2 graduate students in academic writing, the research data reveals their use of code glosses compared to L1 student writers, providing pedagogical implications and academic training possibilities.

2. Literature review

As metadiscourse devices, code glosses “help to contribute to the creation of coherent, reader-friendly prose while conveying the writer’s audience-sensitivity and relationship to the message” (Hyland, 2007, p. 266). Their role is to clarify, elaborate and embellish a plausible argument and, furthermore, these devices tend to clarify the writer’s communicative purpose (Hyland,

2007), which is a prominent feature of English academic writing. Code glosses allow the writer to explain, rephrase, reconsider, summarize, and sometimes even distance themselves from the previously presented information (Bondi, 2020). Reformulations are purposeful, since the author re-elaborates the idea to facilitate reader's comprehension. They can be regarded as an auto-reflection about language (Cuenca & Bach, 2007), but they are also regarded as paraphrasing (Bondi, 2020), or linguistic expressions that establish a link between old and new knowledge. Bondi (2020) even refers to them as repairs of the aforementioned discourse.

Code glosses have two main functions: reformulation and exemplification (Hyland, 2007). Markers, which are specific words or phrases that help organize discourse, clarify meaning, and guide the reader through a text, play a crucial role in both functions. In the case of reformulations, markers (e.g., *for instance, in other words*) are used to restate or elaborate a statement in different words, either to present an alternative perspective or to reinforce the message. According to Hyland (2007), reformulation code glosses can expand or reduce the propositional information, depending on their function in the text. On the one hand, expansion can be achieved by providing an explanation (a clarification of the preceding unit) or an implication (drawing a conclusion or summing up the import of the previous unit). Reduction, on the other hand, is achieved by paraphrase (restating the idea in different words) or specification (providing additional information of the statement). Murillo Ornat (2019) provides a more detailed classification, dividing reformulation markers into categories for identification (helping the reader in reference assignment), specification (helping the reader by including the presence of a cataphoric element), explanation (reformulating and clarifying the issue), definition (providing the implicit or encyclopaedic information to understand a certain term), denomination (providing implicit conceptual information) and conclusion (constituting an implication arising from the initial sequence) with the subtype of mathematical operation (providing a conclusion where a calculation is performed and made explicit). In the same paper, (Murillo Ornat, 2019) reformulation markers are also divided in parenthetical sequences, signalled by parentheses, dashes, or commas, where a content is inserted and the rhetoric linearity is not disrupted. A similar pattern was used by Dafouz-Milne (2008), who added the categories of parentheses and punctuation devices to code glosses. In regarding of three minute thesis academic competition and spoken usage of code glosses, Liu et al. (2023) added definitions (a statement to explain

concepts unfamiliar to the non-specific audience) and analogies (creating a parallel from experience and concepts from outside the specialist research) to Hyland's (2007) categories, noticing in their research that these devices were used less than exemplifiers but more than reformulators.

Exemplifications, on the other hand, are used to clarify or support a certain meaning by providing an example, which can be a group, clause, word, or numerical figure (Hyland, 2007). Exemplifications are common communicative devices since they help authors make their ideas accessible and persuasive. These examples may be used to clarify general questions, to offer specific illustrations of general phenomena, and thus make them closer to and more easily remembered by the audience (Calsamiglia & Van Dijk, 2004). By using examples, the author indicates that they have certain presuppositions related to the readers' familiarity with the topic and with the world knowledge in general (Dehghan & Chalak, 2016). Hyland (2007) distinguishes exemplification markers that offer an instance of a general category, provide a parallel, or provide a rule.

However, there seems to be a gap in research regarding the use of code glosses as they appear to be only randomly covered in literature. Cuenca and Bach (2007) deal with the form and usage of reformulation markers in academic papers written in English, Spanish and Catalan, concluding that Spanish and Catalan papers have more of them in order to identify expansion, reduction and permutation. Bondi (2020) investigates their use in academic blogs, finding that A-list bloggers use code glosses similar to academic journalists, though they rather use them to engage readers in the development of the text than in instructing them. Bondi's corpus presents a preference for the following code glosses: *say*, *like*, *in fact* and *especially*. Investigating 120 research articles in three soft disciplines, Cao and Hu (2014) found elaboration and delimitation reformulators to be the most common. The former implies the second unit that elaborates or clarifies the first unit, while the latter refers to restatements in the second unit that demarcate the scope of propositional material in the first discourse unit.

In a comparison of native and non-native writers, Dehghan and Chalak (2016) found that there is no significant difference between the frequency of code glosses in introduction sections of native and Iranian non-native English academic authors, which is also relevant to the present study. Their conclusions reveal that reformulations are used in approximately 55% and exemplifications in 45% of code glosses in introductions, which are used by

authors to present complex ideas by elaborating them, so the use of examples is not common in this section. Edusei (2015) also found a higher frequency of reformulation markers in her corpus of BA theses of non-native speakers of English in Albania and Germany, though her corpus exhibits significant differences in the use of code glosses between these two groups of students. Guziurová (2020) compared 72 first draft research articles from ten different L2 background authors with L1 English authors, concluding that, when entire research articles are analysed, exemplification markers are used slightly more in both corpora than reformulation markers. It is Hyland's (2007) conclusion as well, since soft disciplines need contextualization and authors cannot assume readers' background knowledge. Using the same SciELF corpus as Guziurová (2020), but adding both soft and hard disciplines in the research, Murillo Ornat (2019) noticed similar frequencies of the use of code glosses; however, the author could not single out any trending patterns in their use, claiming it to be a true melting pot of usage.

In Guziurová's (2020) research, non-native writers tend to use more reformulations than native ones (such as *this is*, *in other words*, *namely*), while the latter used more examples (such as *such as*, *e.g.*, *for example*, *for instance*). Similarly, in Guziurová's (2022) research on master's theses, one of the conclusions is that MA students tend to overuse code glosses, especially reformulation markers, with *i.e.* and *such as* being the most frequent. One of the main purposes for the use of reformulation markers is the simplification by non-native authors, especially since students rely on simple grammatical forms. This general overuse of reformulation markers can also be attributed to a lower level of writing expertise as well as cultural differences that rely on Czech academic writing conventions. In the Turkish context, where MA and/or PhD curricula offer limited academic writing training (Kafes, 2022), experienced writers use code glosses three times more than novice writers, who do not elaborate their ideas to a greater extent and rely mainly on reformulation, ignoring the readers' background knowledge about the issues in question.

Interesting research was presented by Liu et al. (2023), who examined the use of code glosses as a communication strategy in an academic competition with the goal of presenting research by graduate students to non-specialist audience. In this spoken communication, students used examples, definitions, reformulations and analogies to make their research more accessible, coherent and engaging to audience. Additionally, Nasau Triana (2019) researched the strategies students used when having to express

content originally written in essays in their oral presentations, concluding that students mostly relied on modality and code glosses and high-achievers consistently used code glosses.

Previous research on code glosses in academic writing has primarily employed corpus-based quantitative methods, analyzing linguistic features through frequency counts and statistical comparisons (e.g., Hyland, 2007; Guziurová, 2022; Kafes, 2022). Many studies have examined L1 and L2 differences in code gloss use (Dehghan & Chalak, 2016; Guziurová, 2020), while others have explored novice vs. expert academic writers (Edusei, 2015; Kafes, 2022) or cross-linguistic variations (Cuenca & Bach, 2007). Our study builds upon this body of work by adopting a similar corpus-based approach, but extends prior research in several key ways. First, unlike studies focusing on expert academic writing, we aimed to eliminate the possibility of a ‘genre mismatch’, as suggested by Walková (2020), by comparing L2 student writers to L1 student writers rather than L1 expert writers. Second, while previous research often examines shorter academic texts such as essays (Khazaei et al., 2020), our study analyzes MA theses, more specifically the Discussion section of the theses, providing insights into how L2 writers navigate metadiscourse in more extended and complex academic writing. Third, while findings from Spanish, Catalan, Turkish and Czech L2 academic writing suggest a tendency toward reformulation over exemplification (Cuenca & Bach, 2007; Guziurová, 2022; Kafes, 2022), our study introduces Serbian and Croatian L2 writers into the discussion, contributing a new linguistic and cultural perspective. With all this in mind, this paper will try to continue the line of research by attempting to answer the following research questions:

1. What are the most striking frequency patterns of code glosses used by MA level L2 writers from Novi Sad, from Osijek and L1 MA writers?
2. What features are most statistically underused or overused by MA level L2 writers from Novi Sad and Osijek, as compared to the L1 group?

3. Method

The current study presents a contrastive look on code glosses in master’s theses written in English as a foreign language by Serbian and Croatian

graduate writers, on the one hand, and American and British graduate writers, on the other. The aim of the study is to explore how these markers are distributed according to crosscultural and crosslinguistic preferences, with a special focus on the code markers that L2 writers tend to overuse and underuse.

3.1. Corpus

The research was conducted on a corpus of unpublished master's theses in the field of linguistics and applied linguistics written by L2 student writers from Serbia (Faculty of Philosophy, University of Novi Sad) and Croatia (Faculty of Humanities and Social Sciences, Josip Juraj Strossmayer, University of Osijek) and L1 student writers from the United Kingdom (University of Birmingham, University of York and University of Edinburgh) and the United States (Brigham Young University, City University of New York, University of California, University of North Dakota and Portland State University). The corpus consisted of the Discussion chapters from the students' theses as found in the IMRAD structure academic writing, with a total of 646,659 words across three subcorpora (see Table 1). The L2 corpus was compiled from first drafts of MA theses submitted by students at the University of Novi Sad and the University of Osijek. These drafts were collected with the explicit consent of the students, ensuring that only original, pre-revision versions were included for analysis. The L1 corpus was assembled from unpublished theses publicly available in online institutional repositories of universities in the UK and US.

Subcorpora of master's theses, chapter <i>Discussion</i>	Word count
Faculty of Philosophy, University of Novi Sad, Serbia	255,451
Faculty of Humanities and Social Sciences, Josip Juraj Strossmayer University of Osijek, Croatia	121,170
Universities from the United Kingdom and the United States	270,038

Table 1. The distribution of word counts across the three subcorpora.

The corpus, in total, included 376,621 words written by non-native speakers (L2 writers) and 270,038 words written by native speakers (L1 writers).

In order to ensure that the theses were of comparable genres and topics, both L1 and L2 theses were in the fields of linguistics or applied linguistics. This study aimed to eliminate the possibility of a 'genre mismatch,' as

suggested by Walková (2020), by comparing L2 student writers to L1 student writers rather than L1 expert writers. Furthermore, the research reported in this paper centers on the Discussion chapters of the theses, which are often the most intricate in terms of structure and concepts (Bogdanović & Topalov, 2022), relying on the assumption that the abundance of ideas and the length of these chapters would encourage student writers to employ metadiscourse more frequently.

3.2. Token identification and analysis

The primary unit of analysis in this investigation, i.e. the token, consisted of individual words (e.g., *specifically*) or phrases (e.g., *in fact*) serving as code glosses, following the model proposed by Hyland (2005, 2007). Code glosses encompassed both conventional reformulation markers with grammaticalized forms, such as *i.e.*, *that is*, *namely*, *in other words*, and a range of lexical expressions employed to rephrase or elucidate the preceding information, such as *this means* and *as a matter of fact* (Guziurová, 2022).

The process of identifying the tokens involved a two-step procedure. In the first step, the authors utilized Hyland's (2005) comprehensive list of code glosses to locate all instances of their use in master's theses. The code glosses were then extracted using AntConc, a freely available corpus analysis toolkit, and prepared for analysis within their respective sentence contexts. During the second step, the authors manually filtered out from the corpus those instances of the targeted words and phrases that were not employed for metadiscoursal purposes. The remaining glosses were then categorized, which resulted in a total of 1,766 instances of code glosses across all three subcorpora, with 775 code glosses used by L1 student writers and 991 code glosses used by L2 student writers. Of these, 687 instances were written by students from the Novi Sad subsample and 304 by students from the Osijek subsample.

The absolute frequency was established by recording each occurrence of an item in a corpus. Subsequently, relative frequencies were computed, with the results normalized to 10,000 words. In studies that contrast corpora of varying sizes, normalization prevents discrepancies that arise from different word counts, as longer texts naturally contain more occurrences of any given linguistic feature. The normalization process involved three steps: 1) first, the total number of times each code gloss appeared in a corpus (e.g., the number of times *for example* occurred) was recorded, 2) next, this raw

frequency was divided by the total number of words in the corpus and multiplied by 10,000 ensuring that frequencies are presented relative to a fixed word count, allowing for fair comparisons, following which 3) the resulting normalized values allowed for direct comparison between L1 and L2 writers, highlighting patterns of overuse and underuse. To assess statistical variations in frequency counts between L1 and L2 student writers, log-likelihood tests were employed. The Log Likelihood Wizard, developed by Paul Rayson at the University Centre for Computer Corpus Research on Language, Lancaster University, was used for this analysis.

4. Results

The following section presents a comprehensive analysis of the use of code glosses in the unpublished MA theses in the field of linguistics and applied linguistics of L2 student writers from Novi Sad and Osijek, contrasted with L1 English student writers from the UK and the USA. As outlined in the methodology section, our study normalized results per 10,000 words to quantify the prevalence of exemplification and reformulation (see Table 2).

	NS	OS	L1
reformulation	15.31	15.10	10.92
exemplification	11.59	9.99	17.78

Note: NS corresponds to Novi Sad; OS corresponds to Osijek.

Table 2. Code gloss markers by subcorpus (per 10,000 words).

As evident from the density of code glosses, Novi Sad and Osijek L2 student writers both have higher word counts for reformulation compared to L1 English novice writers, with Novi Sad having a slightly higher count than Osijek. L1 English writers have a significantly higher word count for exemplification compared to L2 student writers both from Novi Sad and Osijek.

4.1. Reformulations

Table 3 displays total percentages of reformulations found in MA students’ theses across the three subcorpora, in the descending order of frequency of use.

NS (%)		OS (%)		L1 (%)	
i.e.	23.13	that/this/which means	31.8	in particular/particularly	29.1
namely	19.95	i.e.	29.2	i.e.	11.1
that/this/which means	17.46	namely	16.7	in fact	9.1
in particular/particularly	11.34	in other words	6.8	that/this/which means	8.9
in other words	9.52	in particular/particularly	4.7	namely	8.7
that is to say	6.8	in fact	4.2	in other words	8.2
in fact	3.63	specifically	2.6	indeed	7.2
indeed	3.17	that is to say	2.1	that is to say	6.3
specifically	2.95	indeed	0.5	specifically	5
defined as	1.13	defined as	0.5	that is	3.6
known as	0.45	known as	0.5	defined as	1.7
that is	0.23	I mean	0.5	known as	1
as a matter of fact	0.23	that is	0	I mean	0.2
I mean	0	as a matter of fact	0	as a matter of fact	0
TOTAL	100	TOTAL	100	TOTAL	100

Table 3. Percentages of reformulations across the subcorpora.

The overall distribution of code glosses used by Novi Sad and Osijek novice writers indicates a preference for a smaller repertoire of reformulations, with the five most frequent glosses (*i.e.*, *namely*, *that/this/which means*, *in particular/particularly* and *in other words*) accounting for the greatest percentage of use (81.4% for the Novi Sad subcorpus and 89.2% for the Osijek subcorpus, compared to 66.9% for the L1 subcorpus). There is also a notable concentration on the expressions *that/this/which means* and *i.e.* found among Osijek student writers, indicating a preference for a pair of reformulation strategies which can account for roughly two thirds of total uses of reformulations. The percentages for L1 student writers show a more balanced distribution across different expressions, suggesting a more varied and evenly spread repertoire in the use of reformulation strategies. Novi Sad and Osijek novice writers seem to favor specific expressions to a greater extent, while L1 students demonstrate a more diverse and evenly spread use of reformulation expressions. Typical illustrations of the use of reformulations are given in Examples 1, 2 and 3, with examples showing the most frequently used code gloss for each subcorpus.

- (1) The second participant reported primarily linguistic goals *i.e.* introduction of the new vocabulary, learning of synonyms, expansion of the students' vocabulary and lastly familiarization with a new historical figure. (NS student writer)

- (2) Results have shown that participants are at the intermediate level of ICC, **which means** that they are able to use certain principles to deal with intercultural situations, rather than improvise reactively. (OS student writer)
- (3) The present study was designed to inspect how and to what extent L1 orthographies influence the L2 word recognition. **In particular**, the presupposition of the empirical study assumed that readers usually tend to apply their L1 orthographic depth backgrounds to their L2 reading. (L1 student writer)

4.2. Exemplifications

The total percentages of the use of exemplifications among MA student writers are shown in Table 4.

	NS	OS	L1
such as	62.5	51.2	54.8
for example/for instance	23.6	22.3	32.9
e.g.	12.8	25.6	12.1
say	1.0	0.8	0.2
TOTAL	100	100	100

Table 4. Percentages of exemplifications across the subcorpora.

A comparative look into the displayed percentages reveals similar patterns of use across all three subcorpora, with *such as* being by far the most preferred code gloss used to illustrate the point the writers are trying to make. Typical uses are shown in Examples 4, 5 and 6.

- (4) Participants who work longer as programmers seem to believe that speaking microskills **such as** agreeing and disagreeing in English are more important than programmers who work for a shorter period of time. Also, they believe listening microskills, **such as** the ability to identify the topic and follow its development are also crucial, whereas participants who work shorter believe these microskills are not so important for their work. (NS student writer)
- (5) However, as the level of language learning raises, the attention switches to more cultural and personal aspect, especially with poetry (cf. Table 3); **e.g.** critical discussion, or in other words, developing students' awareness of the literature in the foreign language. (OS student writer)

- (6) Qualifiers occur less frequently in the Times (11 in the total sample and two in the top 30 words), but still often indicate political processes or agents (“state,” “political,” and “royal,” **for instance**); the word “prime” is often used in the phrase “prime minister,” “Independent” refers to a British political party, and “economic” makes an appearance here as well. (L1 student writer)

4.3. Overuse of code glosses by L2 writers

In order to identify significant differences in the overuse and underuse of code glosses by Novi Sad and Osijek student writers compared to L1 student writers, we conducted a series of Log-Likelihood tests, which are statistical tests used to determine whether differences in the frequency of terms between groups are significant. A higher log-likelihood value suggests a stronger difference between the groups. Table 5 shows the code glosses that were statistically significantly overused by L2 student writers compared to L1 student writers.

	NS	OS
Code gloss	Log-Likelihood	Log-Likelihood
i.e.	24.95***	24.95***
namely	25.48***	7.67**
that/this/which means	16.65***	40.5***
reformulation	19.34***	11.51**

Note: *p<.05, **p<.01, ***p<.001, ****p<.0001.

Table 5. Code glosses overused by L2 writers compared to L1 writers.

In comparison to L1 student writers, both L2 student writers from Novi Sad and L2 student writers from Osijek significantly overuse a total of three code glosses, all from the group of reformulations. The reported Log-Likelihood coefficients indicate that the overuse ranges from slight (*namely* for the Osijek student writers) to highly noted (*i.e.* and *this/that/which means* for both the Novi Sad and the Osijek subcorpora and *namely* for the Novi Sad subcorpus). In terms of the total volume of reformulation code glosses used by Novi Sad and Osijek student writers, there is also a highly pronounced overuse of glosses found in comparison to L1 student writers.

4.4. Underuse of code glosses by L2 writers

As observed in Table 6, in contrast to L1 student writers, both subsamples of L2 student writers underuse code glosses.

	NS	OS
	<i>Log-Likelihood</i>	<i>Log-Likelihood</i>
in fact	-8.05 ^{**}	-4.42 [*]
indeed	-5.1 [*]	-15.75 ^{****}
in particular/particularly	-26.58 ^{****}	-45.37 ^{****}
that is	-13.93 ^{****}	-11.12 ^{***}
for example	-40.73 ^{****}	-17.89 ^{****}
for instance	-30.15 ^{****}	-26.65 ^{****}
such as	-9.66 ^{***}	-23.54 ^{****}
exemplification	-34.43 ^{****}	-35.79 ^{****}

Note: ^{*} $p < .05$, ^{**} $p < .01$, ^{***} $p < .001$, ^{****} $p < .0001$.

Table 6. Code glosses underused by L2 writers compared to L1 writers.

Slight underuse is found with *in fact* for both Novi Sad and Osijek student writers and with *indeed* and *such as* for Novi Sad student writers. The remaining code glosses are significantly underused by L2 student writers from both universities, with the total use of exemplifications also significantly underused by L2 student writers compared to L1 student writers. The following section will explore the implications of the results, along with the connections to the existing literature and broader significance within the study's scope.

5. Discussion

The aim of this paper was to explore the most prominent frequency patterns of the use of code glosses among MA level L2 writers from Novi Sad and Osijek, on the one hand, and L1 writers from several universities from the UK and the US, on the other. The study further aimed to identify the features that are perceived as underused or overused by MA level L2 writers in comparison to the L1 group. The quantitative analysis of code gloss use in the corpus of the Discussion chapter of MA theses revealed intriguing patterns in reformulation and exemplification strategies. Most notably, L2 writers demonstrated higher word counts for reformulation, with distinct preferences for specific glosses, and an overall overuse of reformulation glosses and underuse of exemplification glosses compared to L1 student writers. The discussion explores these findings, shedding light on the nuanced differences in code gloss usage and their potential implications for academic writing proficiency and cross-cultural communication.

The investigation into code glosses stems from the understanding that these linguistic elements serve as indicators of broader cultural tendencies, uniquely shaped by individual approaches, rather than perceived as errors (Eckstein & Ferris, 2018). While acknowledging the importance of maintaining individual expression, this analysis draws attention to the fact that writers should align with global-level features adhered to by their audience in order to ensure successful reception of their writing.

Overall, the current analysis of code gloss use revealed noteworthy differences in the use of exemplification and reformulation across the studied cohorts. Specifically, the data highlights distinctive preferences in writing styles between L2 and L1 writers, with L1 writers demonstrating a higher inclination towards exemplification, and L2 writers, particularly those from Novi Sad, exhibiting a propensity for reformulation. While the results reveal a comparable density of overall glosses per 10,000 words across the examined L1 and L2 cohorts, the relative frequencies show an imbalance, both in terms of the categories of glosses and among the different subcorpora. The observed results are largely consistent with findings from previous similar studies. Specifically, in a study by Kafes (2022), EFL student writers were also discovered to overuse reformulation code glosses and underuse exemplification glosses compared to L1 student writers, while the overuse of reformulations was also confirmed by Guziurová (2022). Similarly, while investigating expert writers, Hyland (2007) found that the majority of code glosses in the field of applied linguistics were, in fact, employed for exemplification purposes. Our results confirm the observation made by Siepmann (2005) that even writers at advanced stages of language learning, in our case consisting of English majors who are at the final phase of their formal education, experience difficulty in forming and using exemplifiers. This observation also not only echoes Kachru's (2009) view that best academic writing practices can be acquired "through lengthy formal education" (p. 111), but also stresses the need for explicit and planned training in academic writing (Rujević & Popović, 2018). Interestingly, a look into the total count for code glosses, irrespective of their category, returned similar results across the investigated subcorpora, which is in contrast with Kafes' (2022) and Khazae et al.'s (2020) findings. While these studies investigated EFL writers, which is similar to our context, they analyzed shorter academic essays written by learners who were at lower levels of English proficiency compared to our cohort. We believe that both of these factors may account for the differences in the respective findings.

An interesting similarity of results can be drawn between L2 Serbian and Croatian and L2 Spanish graduate writers. Compared to the study by Cuenca & Bach (2007), all participants in theirs and this study overused reformulation markers. Spanish and Catalan students used them to reduce the content of their papers or to reduce the implicature of the previous formulations. However, unlike Serbian and Croatian students who have a preference for a smaller repertoire of reformulations, Spanish and Catalan students exploit a greater variety of reformulation markers than their L1 English colleagues, and also use more complex and more variable reformulators in comparison. Spanish newspaper writers also tend to use code glosses much more frequently than their English counterparts, mostly relying on parentheses and reformulators such as *that is*, *in other words* (Dafouz-Milne, 2008). The marker *that is to say* is reported very frequent among L1 Spanish students studying at a Swedish university (Murillo Ornat, 2019), as well as it is reported rather frequent by L1 Serbian students in this study.

While extensive research has examined metadiscourse use among L2 writers, studies on code glosses have predominantly focused on English as an L1 and L2 within well-researched linguistic contexts such as Spanish, Turkish, Persian, and Czech (Cuenca & Bach, 2007; Kafes, 2022; Khazaei et al., 2020; Guziurová, 2022). Comparatively, Slavic-language L2 English writers, including Serbian and Croatian learners, remain underrepresented in the literature. The findings of our investigation align with previous studies indicating that L2 writers tend to overuse reformulation markers while underusing exemplification (Dehghan & Chalak, 2016; Guziurová, 2022). However, unlike studies that found greater variation in reformulation strategies among Spanish or Catalan writers (Cuenca & Bach, 2007), our results suggest that Serbian and Croatian graduate students rely on a narrower repertoire of reformulation markers, primarily *i.e.*, *namely*, and *that/this/which means*. This preference for a limited range of reformulations may reflect transfer effects from Slavic-language conventions, as well as a lack of explicit training in academic writing. Furthermore, compared to previous findings in Turkish and Persian L2 academic writing (Khazaei et al., 2022; Kafes, 2022), where academic instruction on metadiscourse is reportedly limited, the Serbian and Croatian student subcorpora demonstrate a similar pattern, reinforcing the need for pedagogical interventions focused on expanding students' awareness of exemplification strategies. By integrating Slavic L2 English academic writing into the broader

discussion on metadiscourse, this study highlights the role of linguistic and cultural background in shaping metadiscourse strategies and calls for further cross-linguistic investigations to identify whether these trends extend to other Slavic L2 contexts.

While general scarcity could be observed in the investigations focusing specifically on the use of code glosses in L2 academic writing, the findings of studies investigating broader metadiscourse topics provide a useful foundation for the interpretation and contextualization of our current analysis. These studies identify some of the possible reasons behind the overuse, underuse and misuse of metadiscourse by focusing on cultural and linguistic distinctions, along with divergent norms in academic writing and instructional considerations. For instance, a comparative study examining the use of metadiscourse between Chinese EFL writers and English majors revealed notable differences, underscoring the impact of cultural and instructional factors on metadiscourse patterns (Ruan, 2019). The same conclusion on culture-driven preference leading to the specific use of metadiscourse and code glosses by Spanish newspaper authors is underlined by Dafouz-Milne (2008). Moreover, limited formal training in academic writing for L2 writers, coupled with variations across disciplines, contributes to the overuse use of certain metadiscourse markers, as evidenced in L2 MA theses (cf. Guziurová, 2022; Kafes, 2022). The distinctions between native and non-native writers in metadiscourse use, influenced by culturally diverse conventions, further emphasize the role of cultural factors in L2 writing (Pavičić Takač & Ivezić, 2019). This also suggests that even if L2 writers are exposed to repeated endorsement of metadiscourse as they gain experience in academia, “it may not be easy for them to reconcile the practice with what is favored in their native languages” (Fife, 2018, p. 65). Likewise, an investigation into the academic writing of Iranian EFL learners revealed that the scant use of code glosses may stem from a lack of awareness regarding their functions and significance in academic writing (Khazaei et al., 2022).

We believe that the identified patterns of use of code glosses by Novi Sad and Osijek student writers are likely influenced by a combination of these factors, including inadequate academic writing training, cross-disciplinary differences, cultural and linguistic variances and a lack of awareness regarding the functions and importance of code glosses in academic writing.

6. Conclusion

This study aimed to explore the frequency patterns of code gloss use among graduate-level L2 writers from Novi Sad and Osijek in comparison to L1 graduate writers from several universities from the United Kingdom and the United States. The findings revealed distinctive preferences in the use of exemplification and reformulation strategies, with L2 writers demonstrating a propensity for reformulation, while underusing exemplification. The observed differences in writing styles between L2 and L1 writers underscore the nuanced aspects of code gloss usage and its potential implications for academic writing proficiency and cross-cultural communication. The results also align with findings from previous research, supporting the overuse of reformulation code glosses and the underuse of exemplification glosses by L2 writers compared to L1 writers (e.g. Khazaei et al., 2020; Guziurová, 2022; Kafes, 2022). A possible conclusion, as stated above, is that the use of code glosses by novice L2 writers is under the influence of a complex interplay of factors, which include, but are likely not limited to, inadequate training in academic writing, differences in culture and language, as well as a lack of awareness in understanding the roles and significance of code glosses.

Ultimately, drawing on research results, this study recognizes the practical consideration that instructors, armed with insights into various L1 and L2 tendencies, can enhance their understanding of learners, thereby enabling more effective and targeted instructional guidance. As such, teaching strategies for L2 writers should prioritize establishing a robust framework for analyzing and adhering to context-specific, global-level expectations in academic writing (Barone & Cargile, 2020). This includes cultivating an understanding of the cultural and linguistic nuances that influence writing conventions, ensuring that L2 writers can navigate and adapt to diverse academic contexts. Furthermore, instruction should emphasize the development of an authorial stance marked by precision (Aull & Lancaster, 2014). L2 writers benefit from guidance that hones their ability to express ideas with clarity and exactness, ensuring that their written communication aligns with the precision expected in academic discourse. This involves improving skills in language use, tone and rhetorical strategies to convey ideas effectively and authoritatively within the academic community. In line with these recommendations, the effective instruction in metadiscourse should always be explicit. Fife (2018) stresses the importance of modelling appropriate use of metadiscourse in academic discourse so that students

understand its purpose and function in different contexts and different genres, further advocating for a multimodal approach to instruction, with visual cues complementing textual input. Practitioners also emphasize the benefits of discussions (Nausa Triana, 2019), which aid students in understanding the pragmatic appropriateness and grammatical correctness of code glosses in different contexts, of feedback (Mazidah, 2022), which helps them identify areas for improvement and develop their skills in using code glosses effectively, and contrastive analysis of L1 and L2 academic texts, i.e. data driven language analysis (Khazaee et al., 2022), which aids students in understanding the frequency and function of code glosses in various contexts. Students would also benefit from learning in more detail that reading and writing should be regarded as context-rich, situational, and constructive acts, and not simply an information-exchange process, especially in the Spanish context (Dafouz-Milne, 2008).

The major limitation of this study is that the findings may not be generalizable to all L1 and L2 writers in other linguistic and cultural contexts and for all genres. Variations in educational systems and academic writing conventions across different countries could impact the frequency and patterns of code gloss usage, as has been stated above (e.g. Fife, 2018; Pavičić Takač & Ivezić, 2019; Ruan, 2019). While our findings offer insights into specific linguistic and cultural contexts, caution should be exercised when extrapolating them to broader populations. Acknowledging these limitations, this study hopefully contributes to the understanding of code gloss use in L2 academic writing and highlights the need for further exploration into the multifaceted aspects that shape linguistic choices in this context.

Acknowledgment

This research was supported by the Bilateral scientific research project between the Republic of Croatia and the Republic of Serbia: Hedging, Stance and Engagement in Novice EFL Researchers' Academic Writing (2019-2022).

Article history:
Received 07 February 2024
Received in revised form 26 May 2025
Accepted 01 April 2026

References

- Alghazo, S., Al Salem, M. N., & Alrashdan, I. (2021). Stance and engagement in English and Arabic research article abstracts. *System*, 103, 102681. <https://doi.org/10.1016/j.system.2021.102681>
- Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written communication*, 31(2), 151-183. <https://doi.org/10.1177/0741088314527055>
- Barone, S. M., & Cargile, C. (2020). L1 and L2 writing differences: From understanding to practice. *English Teaching Forum*, 58(3), 10-19.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20. <https://doi.org/10.1016/j.jeap.2010.01.001>
- Bogdanović, V., & Topalov, J. (2022). Transition markers as rhetoric devices in graduate students' writing. In S. Gudurić, J. Dražić, & M. Stefanović (Eds.), *Jezici i kulture u vremenu i prostoru X/2* (pp. 243-254). Filozofski fakultet.
- Bondi, M. (2020). Academics online: Code glosses across research genres and public communication. In G. Maurizio, S. Maci, & M. Sala (Eds.), *Scholarly pathways. Knowledge transfer and knowledge exchange in academia* (pp. 57-82). Peter Lang.
- Calsamiglia, H., & Van Dijk, T. A. (2004). Popularization discourse and knowledge about the genome. *Discourse & Society*, 15(4), 369-389. <https://doi.org/10.1177/0957926504043705>
- Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics*, 66, 15-31. <https://doi.org/10.1016/j.pragma.2014.02.007>
- Cuenca, M. J., & Bach, C. (2007). Contrasting the form and use of reformulation markers. *Discourse Studies*, 9(2), 149-175. <https://doi.org/10.1177/1461445607075347>
- Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40(1), 95-113. <https://doi.org/10.1016/j.pragma.2007.10.003>
- Dehghan, M., & Chalaki, A. (2016). Code glosses in academic writing: The comparison of Iranian and native authors. *Research in English Language Pedagogy*, 3(2), 21-29.
- Eckstein, G., & Ferris, D. (2018). Comparing L1 and L2 texts and writers in first-year composition. *TESOL Quarterly*, 52(1), 137-162. <https://doi.org/10.1002/tesq.376>
- Edusei, J. (2015). Code glosses in student writing: A comparative study of Albanian and German BA theses. In J. Schmied (Ed.), *Academic writing for South Eastern Europe: Practical and theoretical perspectives* (pp. 119-130). Cuvillier Verlag Göttingen.
- Fife, J. (2018). Can I say 'I' in my paper?: Teaching metadiscourse to develop international writers' authority and disciplinary expertise. *Across the Disciplines*, 15(1), 61-70. <https://doi.org/10.37514/ATD-J.2018.15.1.05>
- Guziurová, T. (2020). Discourse reflexivity in written academic English as lingua franca: Code glosses in research articles. *Discourse and Interaction*, 13(2), 36-54. <https://doi.org/10.5817/DI2020-2-36>
- Guziurová, T. (2022). Glossing an argument: Reformulation and exemplification in L2 Master's theses. *Topics in Linguistics*, 23(2), 18-35. <https://doi.org/10.2478/topling-2022-0009>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, 28(2), 266-285. <https://doi.org/10.1093/applin/amm011>
- Hyland, K. (2009). Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal*, 1(1), 5-22. <http://hdl.handle.net/10722/130168>
- Hyland, K., & Jiang, F. K. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, 45, 40-51. <https://doi.org/10.1016/j.esp.2016.09.001>
- Hyland, K., & Jiang, F. K. (2018). "In this paper we suggest": Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, 51, 18-30. <https://doi.org/10.1016/j.esp.2018.02.001>
- Kachru, Y. (2009) Academic writing in world Englishes: The Asian context. In K. Murata, & J. Jenkins (Eds.), *Global Englishes in Asian contexts* (pp. 111-130). Palgrave-Macmillan.
- Kafes, H. (2022). Communicative functions of

- code glosses in academic discourse. *Selçuk Üniversitesi Edebiyat Fakültesi Dergisi*, 48, 53-72. <https://doi.org/10.21497/sefad.1218358>
- Khazaei, H., Maftoon, P., Birjandi, P., & Rezaie Golandouz, G. (2020). Metadiscourse markers in a corpus of learner language: The case of Iranian EFL learners. *Journal of Language and Translation*, 10(3), 17-34. <https://doi.org/10.30495/tlt.2020.678752>
- Khazaei, H., Maftoon, P., Birjandi, P., & Rezaie Golandouz, G. (2022). Fixing the underuse of code glosses in Iranian EFL learners' academic writing: A corpus-based comparison of three pedagogical approaches. *Teaching English Language*, 16(1), 167-201. <https://doi.org/10.22132/TEL.2022.150320>
- Liu, Y., Tang, R., & Lim, F. V. (2023). The use of code glosses in three minute thesis presentations: A comprehensibility strategy. *Journal of English for Academic Purposes*, 63, 101239. <https://doi.org/10.1016/j.jeap.2023.101239>
- Luzón, M. J. (2013). Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience. *Written Communication*, 30(4), 428-457. <https://doi.org/10.1177/0741088313493610>
- Mazidah, E. N. (2022). Code glosses of metadiscourse in students' academic presentation. *Journal of English Teaching, Applied Linguistics and Literatures*, 5(1), 63-68. <http://dx.doi.org/10.20527/jetall.v5i1.11977>
- Murillo Ornat, S. (2019). Reformulation and its markers in unpublished research articles: Some evidence on the rhetorical patterns of written academic EFL. *Revista de Lenguas para Fines Específicos*, 25(2), 26-43. <https://ojsdpdc.ulpgc.es/ojs/index.php/LFE/article/view/1187>
- Nausa Triana, R. A. (2019). Modality and code glosses to transition from academic written to oral discourses: An exploratory study. *Íkala, Revista de Lenguaje y Cultura*, 24(1), 51-67. <https://doi.org/10.17533/udea.ikala.v24n01a02>
- Pavičić Takač, V., & Ivezić, S. V. (2019). Frame markers and coherence in L2 argumentative essays. *Discourse and Interaction*, 12(2), 46-71. <https://doi.org/10.5817/DI2019-2-46>
- Ruan, Z. (2019). Metadiscourse use in L2 student essay writing: A longitudinal cross-contextual comparison. *Chinese Journal of Applied Linguistics*, 42(4), 466-487. <https://doi.org/10.1515/CJAL-2019-0028>
- Rujević, G., & Popović, U. (2018). Analiza uspešnosti spontanog usvajanja veštine pisanja apstrakata na osnovnim akademskim studijama filozofije [Analysing the effectiveness of spontaneous skill acquisition pertaining to writing scientific abstracts among bachelor students of Philosophy]. *Metodički Vidici*, 8(8), 31-49. <https://doi.org/10.19090/mv.2017.8.31-49>
- Safari, I. (2018). A corpus-based contrastive study of code glosses used in English academic articles written by authors of politics and applied linguistics. *International Journal of Linguistics*, 10(2), 40-47. <https://doi.org/10.5296/ijl.v10i2.12242>
- Siepmann, D. (2023). This deserves a brief mention: A multi-corpus comparison of written and spoken academic discourse in English and French with implications for pedagogy and lexicography. *Languages in Contrast*, 23(1), 60-92. <https://doi.org/10.1075/lic.21004.sie>
- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (Vol. 1). University of Michigan Press.
- Walková, M. (2020). Transition markers in EAP textbooks. *Journal of English for Academic Purposes*, 46, 100874. <https://doi.org/10.1016/j.jeap.2020.100874>

Jagoda Topalov is an Associate Professor of English Language and Linguistics at the Department of English, Faculty of Philosophy, University of Novi Sad (Serbia). Her research focuses on English language teaching methodology, applied linguistics, individual differences and affective factors in language learning. She has published in national and international journals, presented at conferences and authored and co-authored several monographs on motivation and language learning.

Vesna Bogdanović is employed as a Full Professor at the Faculty of Technical Sciences, University of Novi Sad, where she teaches ESP and EAP courses. She (co)authored a number of scientific papers and student textbooks. Her interests include the analyses of ESP textbooks, research on diverse categories of metadiscourse, pragmatic aspects of academic English, English as a Medium of Instruction (EMI) and internationalization of higher education institutions.

Biljana Radić-Bojanić is Professor of English linguistics at the Department of English Studies, Faculty of Philosophy, University of Novi Sad. She teaches Academic Skills, Intercultural Communication, Sociolinguistics, Varieties of English and Second Language Acquisition (BA and MA levels) and has published eight books and over 200 papers in Serbian and foreign journals. Her main research interests are academic writing, vocabulary development in SLA, and English as a medium of instruction.

