

## RAISING AWARENESS OF THE DEFINITE ARTICLE IN THE EFL CLASSROOM: A CORPUS LINGUISTICS APPROACH

UDC 371.67:811.111'36  
81'322.2

Marta Veličković<sup>1</sup>, Jelena Danilović Jeremić<sup>2</sup>


<sup>1</sup>English Department, Faculty of Philosophy, University of Niš,

<sup>2</sup>English Department, Faculty of Philology and Arts, University of Kragujevac

ORCID iDs: Marta Veličković

 <https://orcid.org/0000-0001-6039-9373>

Jelena Danilović Jeremić

 <https://orcid.org/0000-0003-3880-3848>

**Abstract.** *To date, corpus linguistics has been used to study a variety of different grammatical structures, including the definite article (e.g. Conrad and Biber 2009; Yoo 2009; Crosthwaite 2019; Gözen and Köroğlu 2022). Since the corpus-based approach appears to be under-researched in the EFL context, the aim of this paper is to explore its potential use with EFL learners. This approach will be offered as an alternative to what can be referred to as the 'traditional' approach to definite article instruction, which focuses extensively on lists, rules, memorization, and exceptions. By relying on Conrad and Biber's findings (2009), we will explore how high school EFL coursebook content can be used to develop practical, awareness-raising activities that center around the most frequent uses of the definite article.*

**Key words:** *corpus linguistics, the definite article, teaching materials, EFL acquisition*

### 1. INTRODUCTION

One of the reasons why articles are such an important field of study in the ESL/EFL environment is that they are among the most frequently occurring words in the English language. The definite article in particular is in fact the most frequently occurring individual word of the English language (Master 1997). This leads us to two points of interest: one brings home the point of how relevant the use of articles is for ensuring no misunderstandings or miscommunications take place, and the other poses the question of why, despite the clear frequency of exposure, L2 English article production is not at a higher level of accuracy. There is a further comment to be made pertaining to the frequency of occurrence of parts of

---

Submitted April 10, 2024; Accepted May 8, 2024

**Corresponding author:** Marta Veličković

University of Niš

E-mail: [marta.velickovic@filfak.ni.ac.rs](mailto:marta.velickovic@filfak.ni.ac.rs)

speech of the English language, which is that, in academic speech, the noun is the most frequently occurring category (Conrad and Biber 2009). Despite this, errors in article suppliance persist until later stages of proficiency (Ionin and Montrul 2009). Such a situation provides us with the opportunity to consider the potential implications for the EFL process and whether they could lead to any innovations.

The existing, widespread and usual approach to instructing English language learners (ELLs) on article suppliance is referred to as the ‘standard approach’, which by its nature is quite formal. Its main features include the following: considerable emphasis is placed on the morpho-syntactic features of nouns (whether they are atomic or not, singular or plural, countable or uncountable), and its focus is on the three criteria which regulate the occurrence of the definite article, and which are mostly hyperonymous in the overall hierarchy of the definite article use: uniqueness, identifiability, and familiarity. This approach, though mainly favored by coursebook designers the world over, unfortunately lacks any sort of reference to contextualization, i.e. how context affects the use of articles. The emphasis in the standard approach is mostly on anaphora, and almost not at all on cataphora, or the requirement to focus on referential relations in discourse which includes taking lexical items such as synonyms into consideration when choosing whether or not to use the definite article; no reference is made to the discourse space involving the two interlocutors; very little attention is paid to the associative context which allows us to link various referents which, even though they are not explicitly and directly mentioned, can still necessitate the use of the definite article; and finally the lack of reference to naturally occurring language which is even reflected in the somewhat artificial nature of the individual examples of article use for which this approach is known. In sum, without referring to ongoing discourse, where referential relations span across sentences, and without the exclusion of single-sentence examples, as EFL instructors we will continue to find ourselves in a situation that is not favorable for accommodating all the uses of the definite article.

In our opinion, there are two volumes that provide the most adequate illustration of the standard approach: Lyons’s *Definiteness* (1999) and Huddleston and Pullum’s *Cambridge Grammar of the English Language* (2002). Published a few years apart, the classifications of the uses of the definite article are quite similar, consisting of several overlaps. An overview has been provided in the following table, along with the appropriate examples (Table 1):

**Table 1** An overview of the classification of the definite article use according to the standard approach

	Lyons (1999)	Huddleston and Pullum (2002)	Example
	Situational use	Sensory features	<i>Pass me the torch.</i>
	General knowledge	Non-linguistic knowledge	<i>The Prime Minister has resigned.</i>
	Anaphoric use	Prior mention	<i>I have a brown bag and a gray one. The gray one is my favorite.</i>
Accounts of the definite article use	Associative use	Through association (with the object-referent)	<i>I don’t know what the problem is, but I think it has to do with the keyboard.</i>
	Preceding-following information	Modifier relative clause	<i>I know the person who committed that crime.</i>
	Anticipatory-anaphoric use	Predication property	<i>The responsibility of every parent is to keep their children safe from harm.</i>

However, it is not merely in general reference works that we find the aforementioned shortcomings of the standard approach to article instruction. Something similar can be noted in the coursebooks used in the L1 Serbian linguistic environment (but is by no means restricted solely to it). Coursebooks which have to date been subjected to analysis have been found to provide infrequent reference to theoretical accounts of (in)definite article use (Veličković 2021; Danilović Jeremić and Veličković 2024). The methods of instruction which were included provide no explicit instruction, leaving it up to EFL learners to implicitly deduce why a particular article was provided in a particular sentence, or to memorize instances of occurrence by heart, which leads to overgeneralization, article omission and substitution, and potentially miscomprehension (Huong, 2005; White, 2010). In the classification of definite article use, prior/second mention is the one that receives slightly more attention in the EFL coursebooks in Serbia, the primary one being the use with superlatives. These findings are based on an analysis of 20 elementary school EFL textbooks for grades 5 through 8, all of which had been approved by the Serbian Ministry of Education (Veličković 2021). The findings also showed that these coursebooks include few explicit explanations of how articles are used; provide illustrations for article use with (un)countable nouns, superlatives, generic and specific meaning only; contain an insufficient number of explanations, which were found to taper off at higher levels of proficiency; offer frequently uncontextualized examples, listed in the form of single sentences; and do not provide any instruction on how to account for definite article uses such as *Have you heard the news?*.

## 2. A CORPUS LINGUISTICS APPROACH TO ARTICLE INSTRUCTION: FINDINGS AND IMPLICATIONS

The initial findings on which we based this analysis indicate the possibility of applying a frequency-based model of learning, which is specific to the field of Cognitive Linguistics, to article instruction. The advantages of such an approach were outlined by Epstein (2001) and a decade later by White (2010). These two authors proposed approaches that take the discourse space into consideration, contextualize article use, incorporate mental spaces, Idealized Cognitive Models, and by extension semantic frames and figure and ground, and expand on the existing morpho-syntactic classification of the standard approach. Epstein (2001) proposed the addition of three supplementary uses of the definite article, which although much fewer in number, were proposed as they accounted for an increasingly greater number of definite article uses than did the standard approach: discourse prominence, shifts in point of view, and the role/value distinction. The effects of basing an article-instruction program on just these three uses provided quantifiable effects, with a 37.6% increase in accurate article suppliance (Veličković 2017), which unfortunately showed no long-term effects (Rohrer and Pashler 2007; Veličković 2017). Since some improvement was made, the conclusion to be drawn is that additional changes in the instructional process were further needed.

The possibility of looking for another approach to article instruction, one that would be used independently or in conjunction with the cognitive-linguistic one, was considered and analyzed in more detail by the authors (Veličković and Danilović Jeremić 2023). There is indication that naturally occurring language, as exemplified and compiled in current relevant corpora (the BNC, COCA, Longman Corpus Network, inter alia) might hold the key. By

definition, corpus linguistics focuses on authentic language use across a variety of registers (everyday conversation, fiction, newspapers, academic writing), providing insight into how native speakers use English in a variety of situations. There are a number of studies which focused on the impact of corpus linguistics on the EFL teaching and learning process. For instance, Tejada et al. (2015) worked on the CLEC corpus (the CEFR-Labeled English Corpus) which comprises some 200,000 words of the English language. The material extracted from the corpus was used for L2 teaching activities at the English Department where they worked and introduced into the curriculum as a teaching innovation. Furthermore, Lee and Webster (2012) compiled their own corpus which they further analyzed, one composed of samples of second language writing. More recently, Rizvić-Eminović and Hadžić (2021) relied on the COCA and BNC as sources from which to extract material for teaching English phrasal verbs and idioms, while Gözen and Köroğlu (2022) studied (in)definite article use in a corpus consisting of doctoral dissertations, implementing the Key Word in Context (KWIC) concordance program to analyze it.

With regard to the definite article, the corpus linguistics approach stresses the importance of only four of its uses, which are presented as deserving of special attention in the L2 classroom. For example, the prior/second mention (also known as direct familiarity) accounts for only  $\frac{1}{4}$  of the definite article use in spoken and written English (Conrad and Biber 2009). Furthermore, shared context (the immediate environment) as a source of determining article use was found to be most common in conversations. The presence of modification was most frequently noted in informational writing, while inference (the use of synonyms) is most common in writing in general. When it comes to fiction in general, assumed familiarity (or the aforementioned discourse prominence) occurs most frequently. Finally, other reasons include idioms and generic reference. An overview can be found in the following table (Table 2).

**Table 2** The list of the most common reasons for the use of the definite article in conversation and different types of writing (Conrad and Biber 2009, 65)

	Conversation	Fiction	Informational writing
Introduced previously in text	25%	30%	25-30%
Shared situational context	55%	10%	10%
Modifiers of the noun	5%	15%	30-40%
Inference	5%	10%	15%
Other	10%	35%	10%

Further work on a similar topic was conducted by Yoo (2009). He provided a comparative analysis of 21 ESL/EFL grammar books on the one hand, and the corpus findings published in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999) on the other. Some of the findings have indicated that teaching accurate article usage was not foregrounded in most of the grammar books, especially those targeted at ELLs with lower levels of proficiency, and was in a way ‘sacrificed’ for greater competence in spoken communication. The aforementioned grammar books present seven situations which are considered definite (Yoo 2009, 270) and which align with the types of usage of the definite article outlined in the existing literature: second mention (anaphoric use), shared knowledge (which subsumes associative use), situational use, postmodification (cataphoric use), unique items, unique adjectives, and ranking adjectives or superlatives (unique reference). Similar to Conrad and Biber (2009), Biber et al. (1999) take into account four registers: conversation,

fiction, newspaper language, and academic prose, but also the following uses of the definite article, which are more or less congruent with those found in the grammar books: anaphoric, indirect anaphoric (associative use and shared knowledge), situational, cataphoric, and generic use. What was concluded is that situational use is “by far the most common definite article usage in conversations” (Yoo 2009, 273), which was found in only four of the nine grammar books aimed at lower proficiency ELLs, wherein second mention is discussed in detail. The impression that is created is that second mention is the most frequently occurring use of the definite article, which is actually only the case in fiction. Additionally, the grammar books made no reference to cataphora, while the corpus findings indicated that postmodification requires further analyses, as it is the most frequent use in academic prose and newspaper language.

Despite these findings that are readily available from a corpus-linguistic, data-driven approach, in the case of L1 Serbian students of English, who have been found to be susceptible to article omission and article substitution (Veličković 2017), the coursebooks that they are provided with offer very little information which is congruent with these findings. A case in favor of ensuring that article instruction be informed by corpus linguistics can be found in the analysis of elementary school textbooks used in Serbian elementary schools (Veličković 2021). Even though most elementary school students are ranked low in terms of English language proficiency, building a proper foundation is still key. This low level of proficiency is also the reason why textbooks aimed at this population tend to explain ‘how’ articles are used, while glossing over the ‘why’. This unfortunately does not change, irrespective of whether these textbooks were meant for first or for eighth graders: the lack of explicit instruction persists throughout all the series of analyzed textbooks, as does infrequent article use. This generally breeds teacher-dependency among the target population, or alternatively, if the learners are left on their own to draw conclusions, to oversimplification and thus misconceptions.

One of the ways that the corpus linguistics approach could prove useful is that it could be relied on to counteract the impact of countability which is the concept most frequently mentioned in the aforementioned textbooks. For those still wondering why countability does not provide a good foundation for article use, Miller (2005) provides numerous reasons, including L1 and L2 discrepancies regarding countability, as well as the fact that context, or more precisely, intended use, impacts the countability of a noun. However, not only does corpus linguistics build on the existing finding of the definite article being the most frequently used word in the English language, it has also very helpfully provided us with the four most frequent uses of this article, none of which have to do with the (un)countability of the nouns in question. In other words, second mention, shared situational context, modification, and inference are superordinate or overarching terms in relation to countability or most other morpho-syntactic features of nouns.

### 3. A CORPUS LINGUISTICS APPROACH TO ARTICLE INSTRUCTION IN THE EFL CLASSROOM: PRACTICAL APPLICATIONS

What would the implications of this situation be for us as EFL teachers? The response is twofold: as teachers we are required to become tech savvy and to raise awareness of these findings among our population of EFL students. What is expected of us is to become facilitators in the learning process, assisting our learners when interpreting and

analyzing the data they are exposed to (Rizvić-Eminović and Hadžić 2021). Bennett et al. (2010) concluded that one way of doing so would be to conduct mini research projects with our students and encourage them to work independently, observing data, picking up patterns, and reaching a conclusion. We ourselves would be conducting similar processes. To quote Hughes: “Often the individual teacher who becomes interested in moving beyond intuitions and traditional approaches... is often an IT-literate teacher comfortable with using new technology in front of a class, is often a keen champion of ITC in the language teaching classroom, and enjoys the challenging exploratory nature of students and teachers finding answers together” (2010, 405).

One of the reasons behind the requirement for us to become more tech savvy is that compiling corpora, as quantitative data, requires the use of concordance software to determine any potential patterns (Conrad 2002). The findings from the concordance software can be paired with any theoretical approach (Partington et al. 2013), including the usage-based and frequency-based cognitive-linguistic approach. Identified concordances indicate “real-world instances of language” (Baker 2010, 94), i.e. how native language speakers of English actually use language, and as such we should ensure that our corpus of choice is as representative as possible (McEnery and Wilson 1996). Some corpora provide their own tools that can be used to search for word frequencies, e.g.

- News on the Web
- Global Web-based English (GloWbE)
- Corpus of Contemporary American English (COCA)
- Corpus of Historical American English (COHA)
- The TV Corpus
- The Movie Corpus
- Corpus of American Soap Operas

As far as individual concordance tools are concerned, some are readily available online and have versions that can be downloaded free of charge. A case in point is the WordSmith tool. Other query tools have been outlined by Kaya et al. (2022), including:

- The Sketch Engine for Language Learning (SKELL)
- Collocaid
- Just the Word
- Voyant Tools
- AntConc

There is also the CLARIN infrastructure, which consists of an impressive 75 corpora that can be used in the L2 learning context.

Once we have been familiarized with the tools at our disposal, and have decided on our corpus of choice, the question immediately arises: how can we use the data we obtained from the concordance analyses of key words to solve the problems of article instruction? When it comes to article omission and substitution, we can rely on the aforementioned concordances to determine whether an article (the definite article in particular) should be used, as a search of the key word would indicate the frequency of occurrence of a particular determiner - adjective - noun combination. Considering that there is a particular focus on the use of the definite article with superlatives in the studied textbooks, the concordance analyses would also help determine any sequential links between determiners and adjectives. These analyses allow us to make our way through various types of corpora and thus a variety of genres. The findings indicate that of the four uses of the definite article that

have been singled out for frequency by corpus linguistic findings, only previous mention seems to have a consistent presence across genres.

Furthermore, a corpus analysis approach allows us to manipulate the language material we choose to include in our classrooms. In other words, it encourages us to compile our own corpora that will suit the particular needs of our specific group of students. For this purpose we may consider a number of sources, such as books (including the obligatory reading material already assigned to our particular group of learners and suited to their age; it would even be possible to consider finding existing translations of books that our students are already reading as part of their curriculum), newspaper articles, and even transcripts of recordings of L2 production compiled during role-play exercises. This material would lend itself to error analysis and pattern recognition, especially in the case of the aforementioned article omission and substitution problem. It would also be useful to consider using as many different types of corpora to ensure that the most frequently occurring uses of the definite article can be found in a variety of contexts. Moreover, students can be encouraged to rely on concordances independently when practicing writing (cf. Gaskell and Cobb 2004), once they have been trained how to do so.

One way of incorporating EFL textbooks within the corpus linguistics approach might be the following. Once we have picked out a segment of text that we would like our students to review, for example for a vocabulary exercise, it is possible to then proceed and outline any of the articles found in the text itself. For example, the Headway upper-intermediate textbook contains the following text (Soars, Soars and Hancock 2019, 115):

#### **What is the origin of the @ symbol?**

History suggests that the @ in email addresses, commonly referred to as the ‘at sign’, stemmed from **the** tired hands of medieval monks. During **the** Middle Ages, before the invention of the printing press, every letter of a word had to be painstakingly transcribed by hand for each copy of the book. **The** monks who performed these tedious copying duties, usually in Latin, looked for ways to reduce **the** number of individual strokes for common words. Although the word for ‘at’ in Latin, ‘ad’, is short, it was so common that the monks wished it were even shorter.

In this extract, the target vocabulary items have already been highlighted. We could proceed to do the same for the definite article used in the text (emphasis added). The examples could be used to illustrate prior/second mention (i.e. direct familiarity) as in *the monks*, shared context (i.e. the immediate environment) as in *the Middle Ages*, inference as in *the invention of the printing presses*, and the presence of modification as in *the number of individual strokes for common words*. Alternatively, students could be asked to identify the different uses of the definite article on their own, if provided with a classification beforehand, and then elaborate on this issue in pairs or small groups. If repeated regularly, such explicit practice would certainly increase the students’ metalinguistic awareness.

A similar analysis can be performed on the listening material which nowadays accompanies (almost) every unit in EFL textbooks. For instance, the Headway upper-intermediate textbook includes the following listening task in the section titled *Everyday English: Talking about places* (available as a transcript in Soars, Soars & Hancock 2019, 144), which exemplifies various uses of the definite article (emphasis added):

1. I'll never forget my trip to Giza and getting my first sight of **the** pyramids. They stand on **the** edge of **the** desert, taking you back to the **time** of **the** pharaohs. But they're surprisingly near to modern buildings, too. I kind of expected to be right in **the** middle of nowhere.
2. It's a brilliant city. We did a city tour on **the** first day to get a feel for the place, starting at **the** Reichstag. There's a viewing terrace right at **the** very top of **the** building where you can get incredible views of the surrounding area. Er, we stayed in what was **the** east side of the city. And, even though it's been unified since 1989, there's still a different feel to east and west.
3. We were staying on **the** outskirts of **the** city, so we only had a day for sightseeing. We began at **the** Capitol Building and walked the two miles to **the** Lincoln Memorial at **the** other end of **the** Mall. It's kind of a national park with gardens and lots of important museums, like **the** National Museum of American History.
4. I wouldn't have missed it for the world. It really is one of India's must-see destinations. It's a magnificent white building, right on **the** banks of the river, so you get wonderful reflections in the clear water. We also spent some time in Jaipur - I loved **the** colourful streets in the very heart of **the** Old City.

The four paragraphs specifically outlined here illustrate how important it is to use the associative context when it comes to the use of the definite article. Since this is a listening exercise, the speaker is, in a way, using words to create images of various cities or locations the world over, beginning with Egypt and ending with India. So how do we account for the use of the definite article in these situations? When it comes to oral communication, i.e. conversation, the shared situational context could be invoked at once to account for the definite article use: once the name of the location has been directly (or in some instances indirectly mentioned, as we only have access to segments of the conversation), a shared context is created. Under such circumstances, the use of the definite article actually creates a sense of immediacy while a story is being recounted, or sites described. The listener is given the impression that they are actually present at the location being described. This could provide a very good opportunity to point out to the ELLs the frequency of occurrence of the situational use of the definite article (cf. Yoo 2009), as they are doing the listening exercise.

Listening exercises provide another opportunity for us to approach a corpus in a different way and further enhance not only familiarity with the uses of the definite article, but also to enhance the students' auditory comprehension. Thus, instead of having the students follow along with the transcript, the listening comprehension exercise could be played for the whole class, either in its entirety, or in individual segments, depending on the students' level of proficiency. Once the audio material stops, the students could be asked to retell what it was that they have just heard. While one or some of them are involved in an active recall exercise, the other students would actively be involved in a peer correction exercise. On the one hand, it would be possible to monitor retention of article use, make notes of which definite phrases were reported in the same way, which became indefinite or now contained the zero article, but also make note of which of these inconsistencies the remaining class members caught on to. This kind of exercise would not solely be based on enhancing vocabulary retention, as is usually the case, but also on understanding contextualized uses of the definite article.



Another option that a corpus linguistics approach provides us with is the opportunity to include error analysis more overtly in the classroom, especially with higher levels of proficiency. In that case, EFL instructors can use preexisting contextualized examples of definite article use and make any necessary changes. A case in point would be the following example (Conrad and Biber 2009, 66):

1. In a restaurant.

MELI: Did we pay the bill already?

THANH: Not yet. **Our bill is here on a table.**

2. Sometimes it is not possible to decide when we have a satisfactory answer to a problem, **or even when we have a best answer.**
3. I will talk about **some of problems I would be faced with** if I suddenly had to live in a foreign country.
4. First Interstate reported a \$59.9 million loss for the two previous quarters, **as bank continued to be hurt by bad real estate loans.**

Error correction allows for a comparison between different types of article use at the same time, as well as a comparison between the uses of more than one article (in this case the indefinite and the zero article on the one hand, and the definite article on the other). The four examples included in the error correction exercise include a reference to the shared context/the immediate environment (*a table* instead of *the table*), the presence of modification (*a best answer* instead of *the best answer*), and perhaps inference (*some problems* instead of *the problems*).

Along similar lines, a teacher can create their own corpus of students' writing by collecting written assignments (e.g. compositions, dialogs, etc.) or use existing corpora of learner language, as mentioned previously, to prepare error analysis activities. This can be illustrated by the following examples, extracted from submitted coursework either from Serbian to English translation activities, or the essay writing activities of our tertiary-level Serbian EFL students.<sup>1</sup> The examples are grouped based on the type of error: using the indefinite article instead of the definite one

1. On the other hand, there is **an** opinion that this is not mania but rather insanity.
2. It was a hard decision going to his parents' house [...] with **a** request [for them] to immediately get in contact with him.

using the definite article instead of the indefinite one

3. **The** hot wind was blowing.
- and examples of article omission

4. Nowadays, **Ø** majority of writers tend to invent the truth or tell lies in order for their story to be widely read.
5. These days people usually promise something to someone without asking themselves if they can make that promise come true, for they are unfamiliar with **Ø** main reasons why they should never break their promise.

---

<sup>1</sup> Source: the first author's private corpus of student writing.

In the case of examples 1 and 2, the definite article is required due to the presence of postmodification (the relative clause *that this is not mania but rather insanity* and the non-finite *to immediately get in contact with him*). In example 3, the inference to be drawn is that the author in question meant to convey to us a shared situational context as a part of storytelling. Or potentially, it may be an indication of second mention. However, since this was the introductory sentence to a longer segment of writing, it was in fact no more than a descriptive element, one of many, and the use of the definite article in this instance would not bring about a sense of immediacy but would provide an unnecessary prominence to the referent in question. And finally, in examples 4 and 5 the instances of article omission in both cases are illustrations of the omission on the part of the students to use the definite article, primarily due to modification (in the first instance postmodification in the form of a prepositional phrase, *of writers*, and in the second, premodification in the form of an adjective, *main*).

Last but not least, verbal humor can be beneficial in the EFL learning context because it initiates and maintains learner interest, stimulates social bonding, encourages higher-order thinking, and highlights the saliency of information, thus enhancing its retention and recall (cf. Deneire 1995; Lems 2013). Therefore, we suggest that EFL teachers incorporate puns into their teaching practices as an awareness-raising tool.<sup>2</sup> Given that it is the most frequently occurring word in English, the definite article features prominently in short jokes, such as these (cf. Elliott 2010):

Teacher: Johnny, do you know where **the** pyramids are?

Johnny: No, miss, they must be lost. There was a teacher here yesterday asking **the** same question.

Where do sheep go to get a haircut? **The** baa-baa shop.

What letter is never in **the** alphabet? **The** one that you mail.

**The** wedding was so emotional that even **the** cake was in tiers.

Did you hear about **the** actor who fell through **the** floor? It was just a stage he was going through.

In the aforementioned examples we can identify the four most common uses of the definite article proposed by Conrad and Biber (2009) as well as the situational use and postmodification emphasized by Yoo (2009). For example, in the initial exchange there is a good example of inference (*the pyramids*), as well as shared situational context (*the same question*). The joke referring to the letters is a good illustration of the second mention, even though the same noun has not been used again (*letter - the one*). A shared (situational) context could be relied on to account for definite article use in the final two sentences. In the instance of the wedding and the cake, we find that the speaker is actually creating a shared context for themselves and their audience. The inference is that this is a snippet from an ongoing conversation, since the definite article was used to introduce the noun *wedding*, implying at least one other mention of this noun. Under such circumstances it would be easy to continue with *the cake*, as we can expect, from the shared context, which cake is being referred to. The

---

<sup>2</sup> These can be collected by the teacher or borrowed from sources such as books, newspaper sections, and the like and then compiled into a pun corpus.

situation is similar in the case of the final joke, as the context is constructed by the very opening: *Did you hear about...* a standard introduction to jokes that immediately creates a context and invites the listener to join in and participate. The use of the definite article in *the actor* could be ascribed to modification (*the actor who fell through the floor*), or through shared context. Alternatively, the case could be made that both *the actor* and *the floor* are definite noun phrases precisely due to the aforementioned shared situational context. Finally, *the baa-baa shop* example could be explained through inference, in the sense that by definition, a *baa-baa* (barber) shop is actually defined (explained, identified) by means of modification in the preceding sentence.

#### 4. CONCLUSION

The aim of this paper was to shed light on the relatively under-represented topic of corpus-based article instruction in the EFL setting. As a contemporary approach to teaching English, firmly rooted in large amounts of data, corpus-based instruction enables learners to examine patterns of language use in authentic context as opposed to the traditional, textbook-oriented experience of studying sets of rules and their exceptions, if any (Veličković 2021). One way of doing this is by offering students the opportunity to investigate the possible uses of the definite article in concordance lines in one of the publicly available corpora. Another way of doing this is by adapting the existing texts and exercises in coursebooks to suit the purpose of sensitizing learners to the commonest uses of the definite article, that is those identified by Yoo (2009) and Conrad and Biber (2009). Bearing in mind that article omission and substitution remain a persistent challenge for EFL learners, even at more advanced levels of proficiency, we strongly believe that this issue should be addressed systematically, both implicitly and explicitly. Admittedly, EFL coursebooks of the 21st century provide a variety of texts and exercises, but these do not (explicitly) prioritize the comprehension of article use (cf. Veličković 2021; Danilović Jeremić and Veličković 2024). Hence, it is up to language instructors to modify what they have at their disposal so special attention could be paid to articles. Additionally, we propose that humorous content relating to specific article use be part of the EFL instruction since verbal play can significantly improve the learning experience.

**Acknowledgements:** The first author acknowledges that the paper was prepared as a part of the project *Scientific Findings in English Linguistics and Anglo-American Literature and Culture and Teaching Applications*, conducted at the University of Niš – Faculty of Philosophy (No. 336/1-6-01). The second author gratefully acknowledges the financial support from the Ministry of Science, Technological Development and Innovations of the Republic of Serbia (Contract No. 451-03-65/2024-03/200198).

#### REFERENCES

- Baker, Paul. 2010. *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Bennett, Judith, Syliva Hogarth, Fred Lubben, Bob Campbell, and Alison Robinson. 2010. "Talking Science: The research evidence on the use of small group discussions in science teaching". *International Journal of Science Education* 32(1): 69-95.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Conrad, Susan. 2002. *Corpus Linguistic Approaches for Discourse Analysis*. Cambridge: Cambridge University Press.

- Conrad, Susan and Douglas Biber. 2009. *Real Grammar: A Corpus-Based Approach to English*. New York: Pearson Education.
- Crosthwaite, Peter. 2019. *Data-driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.
- Danilović Jeremić, Jelena and Marta Veličković. 2024. *The missing piece of the puzzle revisited: Articles in secondary school EFL textbooks*. Paper presented at the 26th International Symposium on Theoretical and Applied Linguistics, April 19-21, 2024, Thessaloniki: Aristotle University.
- Deneire, Marc. 1995. "Humor and foreign language teaching". *Humor* 8(3): 285-298.
- Elliott, Rob. 2010. *Laugh-Out-Loud Jokes for Kids*. Grand Rapids: Spire.
- Epstein, Richard. 2001. "The definite article, accessibility, and the construction of discourse referents". *Cognitive Linguistics* 12(4): 333-378.
- Gaskell, Delian and Thomas Cobb. 2004. "Can learners use concordance feedback for writing errors?". *System* 32: 301-319.
- Gözen, Bekir Serter and Zehra Köroğlu. 2022. "How definite are the male and female writers?: A corpus based study on the definite and indefinite article use". *Journal of Language and Linguistic Studies* 18(2): 665-684.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hughes, Rebecca. 2010. "What a corpus tells us about grammar teaching materials". In *The Routledge Handbook of Corpus Linguistics*, 1st edition, edited by Anne O'Keefe and Michael McCarthy, 401-412. London: Routledge.
- Huong, Nguyen T. 2005. *Vietnamese learners mastering English articles*. Unpublished PhD dissertation. Groningen: University of Groningen.
- Ionin, Tania and Silvina Montrul. 2009. "Article use and generic reference: Parallels between L1- and L2-acquisition". In *Second language acquisition of articles: Empirical findings and theoretical implications*, edited by Maria del Pilar García-Mayo and Richard Hawkins, 147-173. Amsterdam: Benjamins.
- Lee, John and Jonathan Webster. 2012. "A corpus of textual revisions in second language writing". In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, edited by Haizhou Li, Chin-Yew Lin, Miles Osborne, Gary G. Lee, and Jong C. Park, 248-252. Jeju Island: Association for Computational Linguistics.
- Lems, Kristin. 2013. "Laughing All the Way: Teaching English Using Puns". *English Teaching Forum* 1: 26-32.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Kaya, Omer Faruk, Kutay Uzun, and Hakan Cangir. 2022. "Using corpora for language teaching and assessment in L2 writing: a narrative review". *Focus on ELT Journal* 4(3): 46-62.
- Master, Peter. 1997. "The English article system: acquisition, function and pedagogy". *System* 25(2): 215-232.
- McEnery, Tony and Andrew Wilson. 1996. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Miller, Julia. 2005. "Most of ESL students have trouble with the articles". *International Education Journal* 5(5): 80-88
- Partington, Alan, Alison Duguid, and Charlotte Taylor. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Rizvić-Eminović, Edvina and Azra Hadžić. 2021. "Using corpora in English language teaching: a brief overview and sample of corpus-based activities". *Zbornik radova Filozofskog fakulteta* 19: 27-35.
- Rohrer, Doug and Hal Pashler. 2007. "Increasing retention without increasing study time". *Current directions in psychological science* 16(4): 183-186.
- Soars, Liz, Soars, John and Paul Hancock. 2019. *Headway Upper-intermediate: Student's Book* (5th edition). Oxford: Oxford University Press.
- Tejada, Angeles Zarco, Carmen Noya Gallardo, Carmen Merino Ferrada, and Isabel Calderon Lopez. 2015. "Building a Corpus of 2L English for Automatic Assessment: the CLEC Corpus". *Procedia - Social and Behavioral Sciences* 198: 515-525.
- Veličković, Marta. 2017. *Upotreba određenog člana u engleskom jeziku u teoriji mentalnih prostora: mogućnost primene u nastavi engleskog jezika kao stranog* [The use of the definite article in the framework of mental spaces theory: possible application in the English foreign language classroom]. Unpublished PhD dissertation. Niš: Faculty of Philosophy.
- Veličković, Marta. 2021. "The missing piece of the puzzle: articles in elementary-school EFL textbooks". *Nasleđe* 48: 351-368.
- Veličković, Marta and Jelena Danilović Jeremić. 2023. *How (Not) To Teach the Definite Article*. Workshop held at the Faculty of Philosophy, University of Niš, April 21, 2023.
- Yoo, Isaiah WonHo. 2009. "The English definite article: What ESL/EFL grammars say and what corpus findings show". *Journal of English for Academic Purposes* 8(4): 267-278.
- White, Benjamin J. 2010. *In search of systematicity: A conceptual framework for the English article system*. Ann Arbor: UMI Dissertation Publishing, ProQuest.

## **PODIZANJE SVESTI O UPOTREBI ODREĐENOG ČLANA U NASTAVI ENGLESKOG JEZIKA KAO STRANOG: KORPUSNI PRISTUP**

*Korpusna lingvistika se do sada bavila proučavanjem različitih gramatičkih struktura, uključujući i gramatičku kategoriju određenog člana (npr. Conrad i Biber 2009; Yoo 2009; Crosthwaite 2019; Gözen i Köroğlu 2022). Međutim, čini se da je primena korpusnog pristupa nedovoljno istražena u kontekstu učenja engleskog jezika kao stranog. Skladno tome, cilj ovog rada bio je da ispita tu mogućnost. Korpusni pristup biće ponuđen kao alternativa tzv. „tradicionalnom“ pristupu podučavanju određenog člana u engleskom jeziku koji se u velikoj meri zasniva na sastavljanju lista pravila (i izuzetaka) za upotrebu člana i učenje istih napamet. Oslanjajući se na nalaze Conrad i Biber (2009), ispitaćemo kako se sadržaj udžbenika koji se koriste u nastavi engleskog jezika kao stranog u srednjim školama može iskoristiti za razvoj praktičnih aktivnosti usmerenih ka podizanju svesti učenika o najučestalijim upotrebama određenog člana u engleskom jeziku.*

*Ključne reči: korpusna lingvistika, određeni član, nastavni materijali, učenje engleskog jezika kao stranog*