

Tamara Tošić*

University of Belgrade

Faculty of Medicine

Belgrade, Serbia

<https://orcid.org/0009-0006-8771-2693>

EFL TEXTBOOKS AND GENDER REPRESENTATION: A MIXED-METHOD CASE STUDY

Abstract

The aim of this paper is to examine the way women and men are represented in two EFL textbooks used in Serbian high schools – *Focus 2* and *Solutions Pre-Intermediate*. In order to thoroughly investigate the topic in question, the study employs both quantitative (frequency counts and collocation analysis) and qualitative (visual aid analysis) research methods. The quantitative data alone shows that women are underrepresented in the corpus. When combined with the qualitative analysis, the data reveals, on the one hand, stereotypical portrayals of both women and men and, on the other, efforts to challenge those stereotypes. The pedagogical implication derived from this small-scale case study is that gender bias should be discussed and problematized by teachers and their students.

Key words: gender representation, English as a Foreign Language (EFL), English Language Teaching (ELT), textbook analysis, corpus linguistics, Serbian high schools

* tamara.tosic@med.bg.ac.rs

1. Introduction

The ubiquity of the English language and its use as an unofficial global *lingua franca* have led to the language penetrating all educational levels and contexts. English is now being taught in elementary schools, high schools, private language schools, at universities, workplaces, and the list goes on. The strong presence of English in education has necessitated the creation of a vast number of educational materials, ranging from textbooks to virtual learning aids. While some countries decide to create their own textbooks, others use international textbooks produced by renowned publishers from English-speaking countries, such as Oxford University Press, Cambridge University Press, Longman, etc.

In this introduction, I will argue that it is vital to adopt a critical approach to the analysis of gender representation in English textbooks. In order to proceed with the argumentation, I will quote the definition of education as stated in Encyclopaedia Britannica: “Education can be thought of as the transmission of the *values* and *accumulated knowledge* of a society” (Huq et al. 2023; emp. mine). The two emphasized terms make this definition, in a way, open – they are utterly culture-specific, which means that education (along with educational materials) differs substantially from one culture to another. Moreover, they allow some notions which are taken for granted in society to permeate the educational process. For instance, Jackson (1990: 33–38) discovered an entire hidden curriculum which was in force in classrooms. Also, since gender issues are an essential part of societal values and accumulated knowledge, it is natural and expected that certain attitudes to gender are built into the educational materials. For that reason, the critical analysis of gender in educational materials is relevant and riveting. What makes the issue even more intriguing is the broad use of international EFL textbooks, which will be the subject of the study at hand. We may question which (or whose) values and accumulated knowledge are found in such textbooks. Bearing the listed arguments in mind, we can safely conclude that an analysis of the way gender is handled in educational materials is of great significance.

Nonetheless, many researchers would not agree that contemporary language textbooks are in need of such an analysis. For example, Bori (2018: 44–46) claims that the representation of women in textbooks has considerably improved over the years and that gender issues should only be considered in conjunction with economic and class issues. On the other

hand, when older studies (e.g. Gupta & Yin 1990; Otlowski 2003; Porreca 1984) are compared with recent ones (e.g. Elmiana 2019; Gouvias & Alexopoulos 2016; Sirén 2018; Vu & Pham 2021), it becomes apparent that changes in terms of gender representation in EFL textbooks have been superficial. The fact that the gender-biased status quo is maintained proves conclusively that stand-alone investigations into how women and men are portrayed in EFL materials such as this one should still be pursued. This also leads to the main hypothesis to be tested in this case study, which is that latent gender bias is present in international EFL textbooks used in Serbia, among other countries.

2. Theoretical framework

Having offered arguments in favor of gender representation research on EFL textbooks, I move on to provide a brief review of previous research which is of crucial importance for the paper at hand. Thus, exemplary studies conducted in Serbia, as well as the works which framed this paper will be presented in more detail below.

In Serbia, attention has been devoted to the critical analysis of gender in elementary and high school readers (Serb. *čitanke*). For instance, Opsenica (2003) compares the official pedagogical aims provided in the elementary school curriculum with the pedagogical aims she notices by analyzing readers for the eighth grade of elementary school. She conducts a qualitative content analysis and concludes that there are certain discrepancies between the two pedagogical aims. In terms of gender, she finds that female characters rarely appear in the texts and, when they do appear, they are portrayed through their familial roles – wives, mothers, etc. In the same vein, Stefanović & Glamočak (2008) carry out a combined quantitative and qualitative analysis of gender representation in elementary school readers. They come to the conclusion that, even though noticeable modernization has taken place (e.g. women are depicted as being active, while violence is discouraged for men), women are still underrepresented and patriarchal beliefs are reinforced. It is essential to consider studies such as the two described above in order to highlight that gender representation analysis can and should be performed on a variety of textbook corpora. Unfortunately, what becomes glaringly obvious is that gender bias is present in a great number of textbooks, regardless of their subject.

Another significant study done in Serbia, which is more closely related to the topic of this paper is a PhD thesis by Vučenović (2021) which is aimed at the analysis of gender ideologies in Italian textbooks. In the part of her dissertation focused on corpus-based analysis, the author uses quantitative and qualitative methods to decipher the attitudes taken to women and men in the chosen textbooks. Her results mirror the results of the two previously mentioned studies – the majority of the textbooks are noticeably gender biased, which is discernible through the unequal numbers of male and female characters and the gender stereotyping present in the dialogues, texts and illustrations. Nonetheless, this dissertation reveals some portrayals of women in professional roles which challenge the common gender stereotypes (e.g. female characters appear as doctors, pilots and directors). Hence, we can deduce that textbooks for foreign languages other than English contain a certain amount of gender bias as well.

Now, I will introduce the works which influenced my paper the most and shaped my approach to the subject of gender representation analysis in EFL textbooks.

The first research item on gender bias in EFL textbooks I encountered was Ghorbani (2009). Her corpus encompassed three global English textbook series used in Iranian language institutes. The main method employed in the study, which the author calls critical content analysis, involves interpreting the frequency counts of multiple phenomena. The author counts the following: female and male characters, women and men appearing in occupational, social and domestic roles, firstness and adjectives referring to women and men. She concludes that gender stereotyping and sexism are detectable in the corpus. With regard to Ghorbani's study, it is important to note that while frequency counts are immensely helpful when working on corpora and can be quite revealing, supplying them with qualitative data (e.g. examples from the corpus) is crucial so as to gain a deeper insight into how the examined phenomena are represented in the given corpus.

Building on the notion of a qualitative interpretation of quantitative data, the next significant work which framed my approach is a book on how corpus analysis tools (such as AntConc and Sketch Engine) can be used to study gender, written by Baker (2014). Every chapter of the book is a piece of linguistic research on gender in its own right, where the author provides detailed practical explanations of various ways of using the

aforementioned tools to examine different gender-related aspects of the corpus at hand. As it is known, corpus analysis software and methods, such as frequency counts, concordancing and collocation analysis, can bring to light aspects of language which cannot be perceived by manually going through the text. Nonetheless, Baker makes a salient point throughout this book: researchers should not stop at the statistical data and consider it to be the final outcome of our study, but we should go a step further and interpret the data in conjunction with the context in which it appears.

Lee (2016) follows in Baker's footsteps in her corpus-based study of gender representation in Japanese English textbooks. Her corpus comprises textbooks from four series published in Japan, where they create their own textbooks instead of using the global ones. She makes use of frequency counts to reveal how many female and male characters appear in the corpus, how many gender-marked and gender-neutral constructions there are, which address titles are used for women and men and who is mentioned first in pairs of gender-marked reference terms. The author implements another method, which combines concordancing and collocation analysis, with the aim of ascertaining which adjectives are used with reference to women and men. Her findings demonstrate that women are still underrepresented and subject to gender bias in EFL textbooks in Japan. The method used in Lee's study, along with the practical instructions offered in Baker (2014), served as the basis for the quantitative part of my research.

The last piece of research contributing to the framework of this paper is a qualitative analysis of gender bias performed on a high school textbook published and used in Japan (Clark 2016). The author carries out a free-form but thorough qualitative analysis of dialogues, texts and illustrations found in the textbook. The results this analysis yields can succinctly be summarized using this quote from the paper:

In general, it is not the quantity of males and females that creates bias; it is the qualitative aspects of their lives as they enact their respective social worlds in words and pictures. They are seen to engage in differing activities, which present them in ways suggesting biased social categorizations. (Clark 2016: 4)

The quote and the entire study illustrate beautifully how significant qualitative data is for a linguistic investigation into gender (and for any linguistic investigation in general). For that reason, in this study, I decided to incorporate a purely qualitative approach to the same data first examined quantitatively.

To sum up what has been said so far, a vast amount of evidence points to the fact that gender representation in EFL textbooks is still a relevant topic worthy of further in-depth research. Thus, the aim of this study is to examine the way women and men are represented in EFL textbooks used in Serbian high schools. Furthermore, I have argued for an approach to corpus analysis which combines quantitative and qualitative methods, with special emphasis on the significance of qualitative data itself and qualitative interpretations of quantitative data. In the next section of this paper, I will present the methodology and the corpus used for the study in greater detail.

3. Methodology and corpus

3.1 Methodology

Firstly, it is crucial to stress that this paper presents a case study on gender representation in EFL textbooks. That means that the conclusions reached in this study apply only to the textbooks included in the corpus. There will be no attempts to extrapolate from the results obtained here since the corpus is not representative of all available EFL textbooks.

Secondly, this is a mixed-method case study, which implies that both quantitative and qualitative methods were employed throughout the research process. Thus, the case study comprises two types of analyses – a quantitative and a qualitative one.

The quantitative corpus-based analysis was performed using #LancsBox (Brezina et al. 2021), a corpus analysis tool developed at Lancaster University. As Baker (2012: 225) explains and illustrates, corpus analysis software such as #LancsBox offers a systematic and data-driven approach to (critical) discourse analysis, which will be adopted here. Three functions of the software were utilized: frequency counts (incorporated into the KWIC feature), collocation analysis (the GraphColl feature) and concordancing (the KWIC feature). Frequencies of gender-marked and gender-neutral common reference terms, relational reference terms, personal pronouns and firstness in pairs of gender-marked reference terms were analyzed. For collocation analysis, gender-marked common reference terms and personal pronouns were taken into consideration. As was elucidated in the Theoretical framework, in this paper, the presentation

of the collected data will not be purely quantitative. Critical qualitative interpretations will be offered for each set of data provided by the #LancsBox software.

The qualitative analysis was more free-form. It was intended as a triangulation method for checking the quantitative results and providing representative examples to corroborate the conclusions. The qualitative segment of the study centered on visual aids found in the textbooks. The significance of pictures in the language learning process has been recognized for a long time (see e.g. Wright 1989). Researchers in the field of language teaching methodology have also highlighted that pictures featured in language textbooks can be subject to interpretation. For instance, Baker (2015) quoted Ansel Adams to illustrate the potential that pictures hold: “A photograph is usually looked at – seldom looked into” (Adams, as cited in Baker 2015). Therefore, the purpose of the qualitative segment of this study is not to look at the visual aids found in textbooks, but to look into them and to highlight the importance of looking into them, in research and in the classroom.

3.2 Corpus

Two textbooks constitute the corpus: *Focus 2* (Kay et al. 2016) and *Solutions Pre-Intermediate* (Falla & Davies 2017). The textbooks were chosen randomly out of those that are approved by the Serbian Institute for Education Advancement and the Ministry of Education. They are both used in the first grade of high school in Serbia. Textbooks at this level were chosen for two reasons. First, in the first grade of high school, students are still quite young (approximately 15 years old); thus, they are still highly impressionable. The potential influence textbooks have on the students' stances towards gender issues should be thoroughly examined. The second reason for opting for these textbooks is the CEFR level they were intended for (A2/B1), which offers enough text for the corpus analysis to be viable.

The corpus has 99,499 tokens. *Focus 2* contributes to that with 46,940 tokens, while *Solutions Pre-Intermediate* has 52,559 tokens (Figure 1). Hence, the corpus is considered to be balanced.

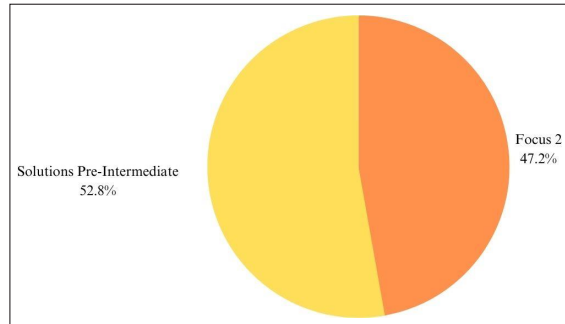


Figure 1. Corpus structure

In order to create a balanced corpus, certain criteria were established. Only the main units of the student's books were included in the corpus. The *Solutions* textbook is significantly longer than *Focus* because of the number of additional materials. Therefore, a decision to exclude the extra materials from both textbooks was made. The goal while building the corpus was to incorporate as much of the text found in the main units as possible since the students are exposed to and possibly influenced by it. So, activity instructions, suggested answers and examples, activities, texts and headings were all included in the corpus. Nevertheless, certain elements had to be eliminated from the corpus. For instance, the listening transcripts did not make it into the corpus. Captions featured on pictures were analyzed separately in the qualitative segment of the study. Numbers, symbols and letters used to enumerate examples, as well as certain activity labels were not included lest they skew the token count. Apart from the listed aspects, everything else found in the main units of the textbooks is a part of the corpus.

4. Results and Analysis

4.1 Frequency counts

4.1.1 *Common gender-marked reference terms*

To start with, the frequencies of common gender-marked reference terms will be presented and analyzed. In the context of this case study, *common reference terms* are nouns which are used to refer to women and men. In

this section, the focus is on gender-marked terms. Therefore, we will see how many times the words *woman*, *man*, *girl* and *boy* appear in the corpus and what the possible implications of those numbers are.

Lemma	Absolute frequency	Relative frequency	Distribution
{man}	90	9.05	17/19
{woman}	31	3.12	16/19
{boy}	49	4.92	13/19
{girl}	72	7.24	14/19

Table 1. Joint corpus

Lemma	Absolute frequency	Relative frequency	Distribution
{man}	32	6.82	8/9
{woman}	16	3.41	8/9
{boy}	30	6.39	7/9
{girl}	49	10.44	7/9

Table 2. *Focus 2*

Lemma	Absolute frequency	Relative frequency	Distribution
{man}	58	11.04	9/10
{woman}	15	2.85	7/10
{boy}	19	3.61	5/10
{girl}	23	4.38	7/10

Table 3. *Solutions Pre-Intermediate*

The first column in Tables 1, 2 and 3 shows the common gender-marked reference terms examined in this section. All of the searches were lemmatized so as to include the cases when the terms appear in their

plural forms in the frequency counts as well. The absolute frequencies (i.e. the total number of occurrences in the corpus) are given in the second column, while the relative frequencies (i.e. the number of occurrences per 10,000 words) can be found in the third column. The distribution column displays the number of units the term appears in. For these frequency counts, the data from the two subcorpora are presented separately since it is particularly important to consider whether one textbook contributes more to the frequency counts than the other.

Bearing previous research in mind, the fact that {man} is the most frequently used common reference term in the corpus does not come as a surprise. Upon closer inspection of the relative frequencies of this term in the joint corpus and the subcorpora, it becomes apparent that {man} is used much more frequently in *Solutions* than in *Focus*. Thus, *Solutions* contributes more to the total of 90 occurrences of {man} in the entire corpus. What does come as a surprise is {girl} being the second most frequently used term in the corpus. *Focus* is the textbook which affects this number more. Be that as it may, there are two possible interpretations of this phenomenon. Since the textbooks are intended for the first grade of high school, the target audience are teenagers. If the target audience are supposed to be *young men and women*, then *women* are underrepresented – {man} is used three times more than {woman} in the entire corpus. Thus, it can be inferred that there is a slightly patronizing attitude towards women in the corpus since they are not portrayed as young women, but as girls. On the other hand, if *teenage boys and girls* are taken to be the target audience, then *boys* are missing. In that case, it can be concluded that boys are expected to be more manly. Having in mind that {man} is more frequent in *Solutions* and {girl} occurs more times in *Focus*, it can be said that each interpretation holds true for one of the textbooks. Hence, the interpretations are two sides of the same coin and are not mutually exclusive.

4.1.2 Gender-marked relational reference terms

Baker (2014: 89) defines relational identification as the situation when a reference term entails a connection to another person – e.g. *wife*, *husband*, *mother*, *father*, etc. In his research, Baker finds that *wife* is used much more frequently than *husband* and deduces that “*Wife* is so frequent because it implicitly refers to a man” (Baker 2014: 92). We will now see whether the data gathered from the corpus in question points to a similar conclusion.

Lemma	Absolute frequency	Relative frequency	Distribution
{father}	9	0.9	5/19
{dad}	24	2.41	12/19
{husband}	1	0.1	1/19
{son}	2	0.2	2/19
{brother}	23	2.31	13/19
{grandfather}	1	0.1	1/19
{grandpa}	2	0.2	2/19
{grandson}	1	0.1	1/19
{uncle}	12	1.21	5/19
{nephew}	1	0.1	1/19

Table 4. Joint corpus – male terms

Lemma	Absolute frequency	Relative frequency	Distribution
{mother}	18	1.81	8/19
{mum}	22	2.21	13/19
{wife}	5	0.5	3/19
{daughter}	4	0.4	2/19
{sister}	29	2.91	14/19
{grandmother}	2	0.2	2/19
{grandma}	1	0.1	1/19
{granddaughter}	0	0	0/19
{aunt}	16	1.61	4/19
{niece}	0	0	0/19

Table 5. Joint corpus – female terms

Tables 4 and 5 show frequency counts for gender-marked relational reference terms in the joint corpus. Since the two subcorpora mirror the joint corpus, the data from the individual textbooks is not provided in this section.

As can be seen from the two tables above, the frequency counts from the joint corpus reinforce Baker's conclusion. *Mother* and *mum* together are used more often than *father* and *dad* (40 times as opposed to 33 times), while *wife* appears more times than *husband* (5 times and 1 time, respectively). Moreover, there are 97 female relational terms in the corpus, while there are 76 male relational terms. Thus, women are more likely to be portrayed in relation to someone else, particularly a man.

When the frequency counts of common and relational gender-marked reference terms are added up, we get the following results: male nominal reference terms are used 215 times, while female ones appear 200 times. That piece of information alone might lead one to conclude that women and men are equally represented in the corpus. However, the in-depth analyses and interpretations offered in this and the previous section prove that the situation is not that straightforward and that many potential biases are at play. This reaffirms the claim made in the Introduction that simple counting (as in the content analysis method) is not sufficient. In order to get the most accurate results, the data needs to be carefully analyzed, compared, contrasted and interpreted critically, while taking the context into consideration at all times.

4.1.3 Common gender-neutral reference terms

Having studied gender-marked reference terms, we now turn to gender-neutral reference terms, more specifically the gender-neutral use of *man*, *person*, *people* and *human*.

Lemma	Absolute frequency	Relative frequency	Distribution
{person}	64	6.43	17/19
{people}	341	34.27	19/19
{human}	7	0.7	4/19

Table 6. Joint corpus

Lemma	Absolute frequency	Relative frequency	Distribution
{person}	30	6.39	9/9
{people}	171	36.43	9/9
{human}	4	0.85	2/9

Table 7. *Focus 2*

Lemma	Absolute frequency	Relative frequency	Distribution
{person}	34	6.47	8/10
{people}	170	32.34	10/10
{human}	3	0.57	2/10

Table 8. *Solutions Pre-Intermediate*

Tables 6, 7 and 8 display the frequency counts for {person}, {people} and {human} in the joint corpus and the subcorpora. The generic use of *man* is not given in the tables since it appears only three times in the entire corpus. It occurs in a text from *Focus 2* about the moon landing: "... US President JF Kennedy promised to put a *man* on the moon...", "That's one small step for *man*, one giant leap for mankind." (Kay et al. 2016: 29; emp. mine). The low number of gender-neutral uses of *man* is deemed positive since such use of the male reference term is considered to be sexist.

Another positive development is the fact that the gender-neutral reference terms occur as frequently as the gender-marked ones – the former are used 412 times and the latter 415 times. Thus, it can be deduced that there is an increasing effort to use gender-neutral reference terms, especially for groups of people.

4.1.4 Gender-marked personal pronouns

In addition to nominal gender-marked reference terms, gender-marked personal pronouns can also be used to investigate gender representation in a text. In particular, they are a good way of "catching" gender-marked references which are not covered by the frequency counts of nominal reference terms.

Pronoun	Absolute frequency	Relative frequency	Distribution
he/him/his	830	83.42	19/19
she/her	458	46.03	19/19

Table 9. Joint corpus

Pronoun	Absolute frequency	Relative frequency	Distribution
he/him/his	425	90.54	9/9
she/her	243	51.77	9/9

Table 10. *Focus 2*

Pronoun	Absolute frequency	Relative frequency	Distribution
he/him/his	405	77.06	10/10
she/her	215	40.91	10/10

Table 11. *Solutions Pre-Intermediate*

The frequency counts for the third-person singular female and male personal pronouns are provided in Tables 9, 10 and 11. There is an obvious discrepancy between the numbers of female and male characters in the joint corpus as well as in the individual textbooks. Thus, we arrive at the conclusion that women are underrepresented in this corpus.

4.1.5 Firstness

The term firstness applies to situations when two gender-marked reference nouns or pronouns are mentioned together, as in “men and women” or “boys and girls”. The choice of which gender comes first in such phrases could be attributed to sexism and gender bias, especially since male firstness has been proven to be much more common (Baker 2014: 92).

The data obtained from this corpus confirms the claim. Overall, there are 37 cases of male firstness (24 in *Focus 2* and 13 in *Solutions Pre-Intermediate*) and 9 cases of female firstness (6 in *Focus 2* and 3 in *Solutions Pre-Intermediate*). Though there are fewer paired references in *Solutions*, the obvious conclusion is that male firstness is more common overall.

When it comes to the interpretation of the data in question, the idiom principle comes into play. To cite Sinclair (1991: 110), “The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.” Phrases analyzed in this part of our study follow the idiom principle – “boys and girls”, “men and women” and even “he/she” are all ready-made chunks in our minds. Nevertheless, the functioning of the idiom principle itself in these coordinate phrases could (and perhaps should) be analyzed in terms of latent societal gender bias.

A captivating instance of the idiom principle at work found in the corpus is the co-occurrence of *aunt* and *uncle*. Contrary to the expected male firstness, in 7 out of 7 cases of *aunt* and *uncle* being mentioned together, the female reference term comes first: “I was staying with my *aunt and uncle* in Jamaica [...]” (Falla & Davies 2017: 50; emp. mine). This finding can be explicated in light of Baker’s (2014: 92) inference that women are more frequently portrayed in relational roles, which was cited in the section focused on the analysis of relational reference terms. Therefore, the only idiomatic chunk in which women consistently come first is one featuring a term which relates women to other members of the family.

4.2 Collocation analysis

Collocation analysis was performed using the GraphColl feature of #LancsBox. Since the tool makes use of various statistical measures, it provides the most relevant results with high-frequency words. For that reason, only the two most frequent common reference terms and personal pronouns were taken into consideration in this segment of the study.

The main aim of this part of the research is to discover discourse (or semantic) prosodies. According to Cheng (2013: 5178), “semantic prosody refers to the attitudinal meaning, often pragmatic, of a lexical item.” The term discourse prosody (Stubbs 1995) is used in order to emphasize the fact that the attitudinal meaning is constructed through discourse. Since this paper is written in the spirit of discourse analysis, the term discourse prosody will be adopted to denote the explained phenomenon.

4.2.1 Common reference terms

As was mentioned in the introduction to this section, only the most frequent common reference terms for each gender were included in the collocation analysis. *Man* and *girl* are the nominal reference terms which appear the most times in the corpus.

The statistical measure used for this segment of research was Mutual Information (MI), which is strength-based, i.e. it shows how strongly the node (in this case the reference term) and the collocates are attracted to each other. The cut-off point was taken to be 3, while the minimal frequency was 5. The range taken into consideration was 5 words to the left and to the right side of the node. Only the collocations from the joint corpus were analyzed since it offers the largest amount of data. In order to account for the most interesting collocates, concordance lines were studied closely, using the KWIC feature.

Position	Collocate	▼ Stat	Freq (coll.)	Freq (corpus)
L	somerton	9.8868425043...	6	7
L	white	8.5242734493...	6	18
R	police	6.8424491300...	8	77
R	next	6.8362172084...	6	58
R	could	6.1457616422...	5	78
L	said	5.8312509810...	5	97
R	had	5.6128099424...	7	158
M	who	5.0648416134...	6	198
R	his	5.0278474114...	7	237
R	was	4.9569509187...	16	569
R	were	4.9313179598...	5	181
R	on	4.6281090730...	12	536
R	why	4.5123006299...	5	242
L	the	4.3507096770...	110	5955
R	she	4.3174216977...	5	277
R	he	4.1008071470...	8	515
L	a	4.0356297125...	40	2694
L	not	3.7446633089...	5	412
L	of	3.7216970548...	18	1507
R	for	3.6762763436...	10	864
M	that	3.6552792706...	6	526
R	in	3.4282265674...	23	2360
R	to	3.1875364718...	21	2546
R	is	3.1205510616...	7	889

Table 12. Collocates of {man} in the joint corpus

Table 12 presents the collocates of the lemma {man} in the joint corpus, sorted according to MI value. It is evident that the majority of the collocates are function words – e.g. modal and auxiliary verbs, pronouns, prepositions, etc. Nonetheless, the three collocates associated most strongly with the node are content words – *Somerton*, *white* and *police*.

Upon taking a closer look at the concordance lines containing these collocates, we see that the phrase *Somerton man* appears only in the *Solutions* textbook, which features a text about an unsolved murder of the Somerton man. This is a good example of how one text can skew the results when working with smaller corpora. For this reason, the distribution of results and concordance lines should always be taken into account. Be that as it may, this collocate, along with others, contributes significantly to the discourse prosody of *man*.

The phrase *white man* appears 6 times in the entire corpus. Almost all of those occurrences are in *Focus 2*. Examples include “But what first brought a *white American man* to this remote part of the world...?” (Kay et al. 2016: 64; emp. mine) and “Rosa Parks was arrested for sitting next to a *white man* on a bus” (Kay, et al. 2016: 100; emp. mine).

Police is another word which is strongly attracted to the reference term *man*. Six out of the eight overall co-occurrences of these two words appear in *Solutions Pre-Intermediate*. The contexts in which the words are used together are, for instance: “Nobody knew who the *man* was. The *police* continued...” (Falla & Davies 2017: 91; emp. mine) and “...a *man* gave *police* a copy of the book...” (Falla & Davies 2017: 91; emp. mine).

All of the collocates listed above play vital roles in the construction of a discourse prosody for the reference term *man*. From the provided data, it can be inferred that men are repeatedly mentioned in contexts focused on crime and (il)legal activities, especially in *Solutions*. Moreover, white men’s race is seen as a feature worthy of emphasis, particularly in *Focus*.

Position	Collocate	▼ Stat	Freq (coll.)	Freq (corpus)
L	toys	9.685513948...	8	14
L	teenage	9.492868665...	5	10
L	boys	9.377392441...	6	13
R	pink	8.860601051...	10	31
L	boy	8.453340727...	9	37
R	colour	7.907906483...	6	36
M	girl	7.907906483...	6	36
R	who	5.670867431...	7	198
L	as	5.518864330...	7	220
M	if	5.293196769...	8	294
R	school	5.153019128...	6	243
L	for	4.737981631...	16	864
R	that	4.623875135...	9	526
R	are	4.468283496...	10	651
R	from	4.121310260...	5	414
R	what	4.069002867...	7	601
R	on	4.011742452...	6	536
L	a	3.955315989...	29	2694
L	the	3.952983336...	64	5955
R	at	3.814797225...	5	512
M	is	3.696829519...	8	889
L	with	3.618985635...	7	821
R	and	3.467892461...	17	2214
L	of	3.105330432...	9	1507

Table 13. Collocates of {girl} in the joint corpus

When it comes to words which have the strongest connection to the lemma {girl}, Table 13 shows many function words (as was the case with {man}), but also some compelling content words – *toys*, *pink*, *colour* and *school*.

The first three collocates can be attributed to a text found in *Focus 2* which problematizes the fact that toys marketed to girls are almost always pink, which is not the case for boys' toys. Though the formatting of the text appears quite stereotypical, with a pink frame and a picture of a little girl wearing pink, the text itself engages with the stereotype critically, which can be deemed positive.

The word *school* occurs 6 times in the vicinity of the reference term *girl*. Out of those 6 occurrences, 5 are found in the *Focus* textbook. Examples are as follows: "...if a *girl* stays in *school*, it makes a big difference to her." (Kay, et al. 2016: 64; emp. mine) or "Ji-min wears exactly the same clothes as the other *girls* in her *school* because school uniform is compulsory" (Kay et al. 2016: 61; emp. mine). The strength of association between the words *school* and *girl* can be justified by the fact that, historically speaking, girls have relatively recently earned the right to mandatory schooling.

Furthermore, in many rural societies, girls are still not being educated. Thus, the presence of girls in schools is something to be emphasized.

Even though the most striking collocates can be accounted for with different explanations, a discernible discourse prosody of *girl* can still be noticed. According to the corpus, girls (perhaps) like the color pink and toys and their education should be highlighted.

4.2.2 Personal pronouns

The parameters in the GraphColl feature of #LancsBox were slightly altered for the collocation analysis of gender-marked personal pronouns. The statistical measure remained the same – MI with a cut-off point of 3 and minimal frequency of 5. Nevertheless, the range was adjusted to 0 words to the left and 1 word to the right of the node. This alteration was made in order to discover and analyze the verbs associated with the pronouns, i.e. the actions performed by the two genders, which constitute a substantial part of the two genders' overall discourse prosodies.

Collocate	▼ Stat	Freq (coll.)	Freq (corpu)
wanted	6.36408667883...	9	21
knew	6.10105227299...	5	14
wants	5.75640397410...	9	32
decided	5.54208504263...	8	33
didn't	5.43447605772...	9	40
wrote	5.38484515030...	5	23
didn't	5.26455093728...	12	60
took	5.26455093728...	6	30
said	5.23449370832...	19	97
thinks	5.20796755538...	5	26
told	5.17144147338...	6	32
doesn't	4.93440235722...	7	44
had	4.80626028222...	23	158
found	4.60919917666...	8	63
also	4.36408667883...	6	56
went	4.28269832616...	8	79
was	4.01915675499...	48	569
could	3.62300505466...	5	78
used	3.55673175677...	6	98
got	3.48494109351...	6	103
give	3.34943985650...	7	132

Table 14. Collocates of *he* (joint corpus)

Collocate	▼ Stat	Freq (coll.)	Freq (corpus)
felt	6.677275401...	6	21
gets	6.636633189...	5	18
wants	6.484630344...	8	32
said	5.692072291...	14	97
has	5.041686792...	16	174
could	4.521156094...	5	78
was	3.917307911...	24	569
had	3.765811989...	6	158

Table 15. Collocates of *she* (joint corpus)

Tables 14 and 15 list the collocates of *he* and *she* in the entire corpus, sorted by MI value. A certain amount of gender bias is identifiable at a glance – many more actions are attributed to male characters than to female ones. When we take a closer look at the tables, we notice that some verbs appear in both tables – e.g. *get*, *want*, *say*, etc. Hence, it is interesting to analyze the actions which are attributed solely to one of the genders. For male characters, those verbs are: *know*, *decide*, *write*, *take*, *think*, *find*, *go*, *use* and *give*. The verb *feel* is the only one associated exclusively with female characters. These findings all contribute to the discourse prosodies being built in this corpus. Namely, men are portrayed as more active, assuming more responsibility – they know, they decide, they think, they give and they take. On the other hand, female characters are more passive, with heavy emphasis placed on their emotions.

4.3 Qualitative analysis

As was explained in the Methodology section, the qualitative part of this research centered on visual aids from the textbooks, more specifically on the social roles of characters found in them. In this part of the paper, pictures from one unit from each textbook will be presented and analyzed as examples of broader trends from the corpus.

The selected unit from *Focus 2* is Unit 6 – Working life. In Figure 2, which is the first one in the unit, both women and men can be seen practicing stereotypical professions. Women are hairdressers, nurses and bankers, while men are builders, plumbers and programmers. This is mitigated by Figure 3, appearing later in the unit and depicting women

in unconventional roles, such as airline pilots and electricians. Therefore, though stereotypes are at play in this corpus when it comes to the characters' professions, efforts to challenge them are present as well.



Figure 2. Stereotypical professions (Taken from Kay et al. 2016: 72–73)

6.3 Listening

Multiple choice
I can identify specific detail in short conversations and monologues.



airline pilot



childminder



electrician

1 In pairs, look at the photos. What kind of person do you have to be to do these jobs? Use the adjectives in the box or your own ideas.

ambitious	brave	caring	clever
energetic	kind	practical	responsible
sensible	serious		

EXAM FOCUS Multiple choice

2 Listen to three short recordings. Choose the correct picture (A, B or C) for each recording.

1 What is the woman's job?

Figure 3. Challenging the stereotypes (Taken from Kay et al. 2016: 75)

The unit chosen from *Solutions Pre-Intermediate* is Unit 8 – Crimes and Criminals. In this unit, women are completely underrepresented. Namely, out of the 26 criminals appearing in this unit's pictures, 22 are male. Figures 4 and 5 show examples of this phenomenon. Thus, based on the qualitative findings from both textbooks, we can infer that men, as well as women, are subject to stereotypes. In this instance, men are stereotypically portrayed as likely to engage in illegal activities, which is in consonance with the analysis of the collocates of the term *man* presented above.

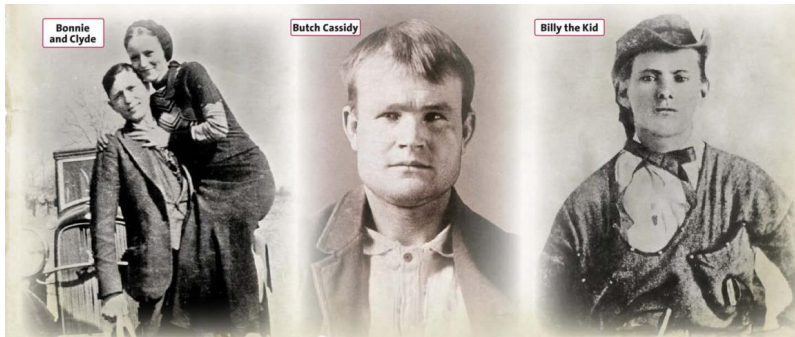


Figure 4. Famous criminals (Taken from Falla & Davies 2017: 87)



Figure 5. Men portrayed as perpetrators of crime (Taken from: Falla & Davies 2017: 92)

5. Discussion and conclusion

The previous section of this paper, offering qualitative data and triangulation, provides a perfect segue into the discussion and conclusion. Bearing in mind the presented findings and analysis, there are several conclusions to be reached. Firstly, as evidenced by the frequency counts, women are underrepresented in the two examined textbooks. Furthermore, the discourse prosodies and the analysis of visual aids show that there are stereotypical portrayals of both women and men. Therefore, the hypothesis given in the Introduction is confirmed when it comes to the corpus at hand. Nonetheless, it must be stated that efforts to combat stereotypes and use gender-neutral terms of reference are apparent, as can be seen from the frequency counts and qualitative data.

Though the studied corpus is rather small, certain pedagogical implications can be drawn from the results. Namely, the gender bias in these textbooks is latent rather than manifest. The bias was forced into the open by the powerful corpus analysis software and the (critical) discourse analysis of the gathered data. Therefore, this paper joins and strengthens the call for a long-awaited reevaluation of EFL materials in terms of gender representation. Bearing in mind that the formal revision of educational materials is a long and intricate process, teachers should take a bottom-up approach and strive to bring the bias to light and present it to the students. Examples of such practice include but are not limited to: problematizing the stereotypical professions assigned to women and men, counting and discussing the numbers of female and male characters in the textbook, considering the contexts in which women and men appear, etc.

Finally, since this study has its limitations (most noticeably the size of the corpus), further research into this topic could attempt to expand the corpus by including different textbook series and levels. Listening transcripts, workbooks and additional materials could also be incorporated into a future study.

References

- Baker, L. (2015). How many words is a picture worth? Integrating visual literacy in language learning with photographs. *English Teaching Forum*, 53(2), 2–13.
- Baker, P. (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. *Critical Discourse Studies*, 9(3), 247–256.
- Baker, P. (2014). *Using Corpora to Analyze Gender*. London and New York: Bloomsbury Publishing.
- Bori, P. (2018). *Language textbooks in the era of neoliberalism*. Oxon and New York: Routledge.
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). #LancsBox v. 6.x. [software package]
- Cheng, W. (2013). Semantic prosody. In C. A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*, Wiley-Blackwell Publishing Ltd, 5178–5184.
- Clark, I. (2016). A qualitative analytic case study of subliminal gender bias in Japanese ELTs. *SAGE Open*, 6(3), 1–8.
- Elmiana, D. S. (2019). Pedagogical representation of visual images in EFL textbooks: A multimodal perspective. *Pedagogy, Culture & Society*, 27(4), 613–628.
- Falla, T. & Davies, P.A. (2017). *Third Edition Solutions Pre-Intermediate: Student's Book*. Oxford: Oxford University Press.
- Ghorbani, L. (2009). *An Investigation of the Manifestation of Sexism in EFL/ESL Textbooks*. ERIC. (10 October 2023) <<https://files.eric.ed.gov/fulltext/ED505434.pdf>>
- Gouvias, D., & Alexopoulos, C. (2016). Sexist stereotypes in the language textbooks of the Greek Primary School: A multidimensional approach. *Gender and Education*, 30(5), 642–662.
- Gupta, A. F. & Yin, A. L. S. (1990). Gender representation in English language textbooks used in the Singapore primary schools. *Language and Education*, 4(1), 29–50.
- Huq, M.S., Browning, R., & Anweiler, O. (2023). Education. *Encyclopaedia Britannica*. (10 October 2023) <<https://www.britannica.com/topic/education>>
- Jackson, P. W. (1990). *Life in Classrooms*. New York and London: Teachers College Press.
- Kay, S., Jones, V., Brayshaw, D., & Michalowski, B. (2016). *Focus 2: Students' Book*. London: Pearson.

- Lee, J. F. K. (2016). Gender representation in Japanese EFL textbooks – A corpus study. *Gender and Education*, 30(3), 379–395.
- Opsenica, J. (2003). Komparacija vaspitnih ciljeva čitanke za osmi razred i vaspitnih ciljeva plana i programa osnovnoškolskog vaspitanja i obrazovanja. *Godišnjak za psihologiju*, 2(2), 165–179.
- Otlowski, M. (2003). Ethnic diversity and gender bias in EFL textbooks. *Asian EFL Journal*, 5(2), 1–15.
- Porreca, K. L. (1984). Sexism in Current ESL Textbooks. *TESOL Quarterly*, 18(4), 705–724.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sirén, T. (2018). *Representations of men and women in English language textbooks: A critical discourse analysis of Open Road 1–7* [Master's Thesis, University of Oulu]. OuluREPO. (10 October 2023) <<https://oulurepo.oulu.fi/bitstream/handle/10024/11321/nbnfioulu-201805312348.pdf?sequence=1>>
- Stefanović, J. & Glamočak, S. (2008). Imaju li čitanke rod? Rod u čitankama u osnovnoj školi. *Philologia*, 6(1), 61–68.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 1–33.
- Vu, M. T., & Pham, T. T. (2021). Still in the shadow of Confucianism? Gender bias in contemporary English textbooks in Vietnam. *Pedagogy, Culture & Society*, 31(3), 477–497.
- Vučenović, N. (2021). *Rodne ideologije u udžbenicima italijanskog jezika kao stranog* [PhD Thesis, University of Belgrade]. NaRDuS. (10 October 2023). <https://nardus.mpn.gov.rs/bitstream/handle/123456789/20791/Disertacija_12699.pdf?sequence=1&isAllowed=y>
- Wright, A. (1989). *Pictures for language learning*. Cambridge: Cambridge University Press.

Received: 20 May 2024

Accepted for publication: 14 October 2024

