

Lexical profile of literary academic articles

Vesna Bratić & Milica Vuković Stamatović

University of Montenegro

vesnabr@ucg.ac.me & vmilica@ucg.ac.me

Abstract

In this paper, we examine the lexical profile of literary academic articles with a view to determining how they differ from research articles in other disciplines and how the vocabulary level and complexity affect reading comprehension, particularly for non-native speakers of English. For this purpose, a corpus of 110 literary articles from reputable journals was compiled and compared against two corpora featuring the same number of articles: one consisting of research articles from Science, Technology and Medicine (STM), and the other comprising research articles from social sciences and other humanities. The results reveal that the lexical profile of literary academic papers is, as expected, more similar to social sciences and other humanities than to the STM field when it comes to the coverage of general-purpose vocabulary, vocabulary level and vocabulary diversity. Despite the lexical similarities to social sciences and other humanities, the vocabulary of literary academic papers is somewhat more complex and diverse than that found in them. The largest differences were noted with respect to the level of academic vocabulary, whose use is much sparser in literary studies than in all other fields. The pedagogical implications include advocating for refraining from reading literary academic articles earlier than postgraduate studies for non-native-speakers of English (with some exceptions), as their vocabulary level will generally be insufficient for those purposes. We also point to the limited value of teaching academic vocabulary to students of literary studies.

Keywords: lexical profiling, literary studies, research articles, vocabulary level, word lists.

Abstrakt

Lexikalisches Profil literaturwissenschaftlicher Artikel

In diesem Referat untersuchen wir das lexikalische Profil literaturwissenschaftlicher Artikel, um festzustellen, inwiefern es von Forschungsartikeln anderer Disziplinen abweicht und wie sich sein Wortschatz und seine Komplexität auf das Leseverständnis auswirkt, insbesondere auf das von Nicht-Englisch-Muttersprachlern. Zu diesem Zweck wurde ein Korpus von 110 literarischen Artikeln aus renommierten Fachzeitschriften zusammengestellt und mit zwei Korpora mit der gleichen Anzahl von Artikeln verglichen: einer mit Forschungsartikeln aus den STM-Wissenschaften und der andere mit Forschungsartikeln aus den Sozial- und Geisteswissenschaften. Die Ergebnisse zeigen, dass das lexikalische Profil von literaturwissenschaftlichen Arbeiten in Bezug auf Erfassung des allgemeinen Zwecks von Wortschatz, Wortschatzniveau und Wortschatzvielfalt erwartungsgemäß eher den Sozial- und Geisteswissenschaften als den STM-Wissenschaften ähnelt. Trotz der Ähnlichkeiten mit den Sozial- und Geisteswissenschaften ist das Vokabular literarisch-akademischer Arbeiten etwas komplexer und vielfältiger als das, worauf in den anderen genannten Wissenschaften gestoßen wurde. Die größten Unterschiede wurden beim akademischen Wortschatz festgestellt, dessen Verwendung in der Literaturwissenschaft viel sparsamer ist als in allen anderen Bereichen. Zu den pädagogischen Auswirkungen gehört, dass es für Nicht-Muttersprachler des Englischen empfohlen wird, keine literarisch-akademischen Artikel vor dem Aufbaustudium zu lesen (mit einigen Ausnahmen), weil ihr Wortschatz für diese Zwecke im Allgemeinen unzureichend wäre. Wir weisen auch auf den begrenzten Wert des Lehrens vom akademischen Wortschatz für die Studenten der Literaturwissenschaften.

Schlüsselwörter: lexikalische Profilerstellung, Literaturwissenschaft, Forschungsartikel, Wortschatz, Wortlisten.

1. Introduction

This study aims at exploring the lexical profile of literary academic papers published in English in literary studies journals. Literary study (or literary criticism) refers to the study, interpretation and evaluation of literature. The discipline is diverse and it includes various schools of thought. Modern literary studies are typically based on literary theory and engage in a philosophical discussion on the methods, goals, genre, style, history, censorship, of literature, on multidisciplinary, etc., as well as on the theory itself, in a clear, cogent and concise way (Tanko, 2017: 43). Academic literary critics generally teach at university literature departments and publish scholarly articles in academic journals. Research articles in literary studies are sometimes labelled as ‘essays’ in some of these journals, which points to

their “liminal status among the other narratives of science” (Stotesbury, 2003: 340). Despite their displaying certain peculiarities in comparison to scholarly articles from other disciplines, in the linguistic literature exploring their features they are thought of as belonging to the genre of research articles and are referred to as such, i.e. as “literary research articles” (Jacoby, 1987; Balocco, 2000; Afros & Schryer, 2009; Shaw, 2009; Lewin & Perpignan, 2012; Malaskova, 2012, etc.). Specifically, Shaw (2009) notes that “research articles in literary studies” typically lack epistemically oriented introductions that follow what Swales (1990) calls a ‘creating a research space’ (CARS) model, but that they nevertheless feature argumentative material that makes claims “just as natural-scientific writing does”, following a pattern: preview claim – quotation – interpretative recount – claim marker – claim. Shaw (2019: 219) concludes that these articles “belong to the broad genre of scientific research articles” and can thus be “compare[d] across disciplines”. We will set out from such a premise in this paper.

In the current linguistic research on the rhetorical structure and language of research articles from various disciplines, the field of literary studies has been relatively neglected (Balocco, 2000; Wilder, 2005; 2012; Tanko 2017). It has also been underexplored within EAP/ESP studies (cf. Lake & Cortes, 2020), i.e. in the research, with the aim of producing pedagogical implications as to the teaching and learning of English for certain purposes and fields. Bearing this in mind, one of the goals of the present paper is to contribute to the research on linguistic variation in research articles from various fields. Our main aim, however, is to ascertain to what extent literary scholarly articles can be used as reading materials by L2 university students in terms of their vocabulary suitability for these students.

To accomplish these goals, we will determine the vocabulary profile of these papers, answering three research questions. First, we will explore how many words are needed to successfully read literary scholarly articles. Next, we will investigate how diverse their vocabulary is. Finally, we will determine how much academic vocabulary they contain. The results obtained will show how similar or different the lexical profile of literary research articles is in comparison with the lexical profile of research papers published in Science, Technology and Medicine (STM), on the one hand, and social sciences and other humanities, on the other hand. We will also determine the vocabulary requirements imposed on non-native English speakers who wish to read these articles.

2. Theoretical background

In the theoretical section of the paper, we briefly present research on the vocabulary load (or level) of various genres and its relation to reading comprehension, modes of measuring lexical diversity, as well as previous studies on academic vocabulary in research papers.

2.1. Vocabulary level

The development of large corpora and software for corpus analysis has allowed for modern lexical profiling studies, which have provided new insights into the lexical frequency of English words in various types of texts. According to their frequency in various genres, as well as in spoken and written discourse, words can be classified into so-called “word lists”. Different lists can be compiled to different purposes: on the one hand, they can be used for teaching and learning a foreign language, which includes research on vocabulary load, course design and testing the knowledge of vocabulary and, on the other hand, they can be utilized for research purposes in the field of psychology (Nation, 2016: 3).

Nation (2013) identifies high-, mid- and low-frequency vocabulary. High-frequency vocabulary includes function and general-purpose words used very frequently in virtually all types of texts; these are generally the words that both native and non-native speakers of English learn first, as they are most exposed to them. The most famous word lists containing such vocabulary typically include about 2,000 words. The first of them in wide use was West’s General Service List (1953), which is now outdated and has been replaced by newer lists, obtained from large corpora, using modern tools and stricter criteria. Two new GSLs (NGSLs) were compiled by Brezina and Gablasova (2013) and Browne, Culligan and Philips (2013), and both report somewhat better results than the original GSL. Some of the word list makers have advocated for these lists to include 3,000 words (Schmitt and Schmitt, 2014).

Mid-frequency vocabulary typically refers to the next 6,000 to 7,000 most frequent words (depending on how the high-frequency range is defined, as either the first 3,000 or the first 2,000 most frequent words). These words are rarer than those classified as high-frequency, but they still make a significant portion of virtually all types of texts.

Nation (2013) designates the rest of the words, i.e. those that are not within

the most frequent 9,000 words, as low-frequency words. They might still be reasonably frequent in some types of texts but they are generally not frequent in everyday texts.

Based on the BNC and the COCA corpora, together containing 450 million words, Nation (2012) produced 25 word lists, each including 1,000 words, sorted according to their frequency in these two corpora. The first two lists generally correspond to the NGSLs mentioned above and can be used as their alternative. The next seven lists correspond to mid-frequency vocabulary, whereas those beyond the ninth list are low-frequency words in general language.

As can be seen, vocabulary level is commonly measured in thousands of words, whereby word itself is differently defined. Most lexical frequency studies equate it with a word family, which means that the base word, together with the forms inflected from it and the words derived from it, fall under the same word family (e.g. *measure, measures, measurement, measurer...*) (Nation, 2013), whereas some studies apply the lemma principle, which is more restrictive and counts only the inflected forms as belonging to the same word (e.g. *measure, measures, measuring...*) (Francis & Kučera, 1982). Nation's word lists employ the word-family principle.

How many words does one need for reading comprehension? This question is somewhat hard to answer but most studies settle for one of the following two thresholds: the first one refers to 95%-coverage of a text, which is enough for “reasonable” or “adequate” comprehension (Laufer, 1989), and the second refers to 98%-coverage of a text (Nation, 2006), which is needed for an “optimal” or “ideal” reading of a text. It is assumed that the rest of the words (2%-5%) can be inferred from the context or using a dictionary etc., without any external support.

Bearing this in mind, a researcher can determine how many thousands of words a reader needs to know in order to reach either adequate or optimal reading comprehension levels of a certain text. For instance, Nation (2006) maintains that either 4,000 or 9,000 word families are needed for reading novels, depending on the threshold we settle for; for newspapers, one needs either 4,000 or 8,000 word families; for children's movies, the figures range from 4,000 to 6,000 word families; and the range for general spoken English varies between 3,000 and 7,000 word families. McQuillan (2016) finds different results for different popular fiction books (some cover 98% with the first 4,000 word families, while some need as many as 8,000 word families

for this level of coverage). Hsu (2011), for her part, determines that 5,000 word families are needed for reading business research articles, if the coverage threshold is set at 95%. Later, Hsu (2013: 465) established that the first 14,000 word families of English covered only 83.1% in her medical textbook corpus (only the lists covering the first 14,000 word families were available at the time of the study), leaving many medical terms outside this scope. Similarly, Vuković-Stamatović (2020) found that the 25,000 most frequent word families of English did not allow meeting the 95%-coverage threshold in a physics research article corpus, implying that a great deal of physics vocabulary comes from low-frequency words which are situated outside these frequency bands. These results point to vast differences in the vocabulary of research articles of various disciplines, which is worth investigating further.

2.2. Lexical diversity

Another aspect of lexical profile is that of vocabulary diversity or variation, i.e. how many different words are used in a text. Vocabulary diversity does not necessarily correspond to vocabulary level. In theory, some texts could contain many rare words, i.e. have a high vocabulary level, but still be repetitive and not display much lexical diversity and, *vice-versa*, a text may display substantial lexical variation but not use many rare words.

Vocabulary diversity can be measured using the so-called *Type-to-Token Ratio* (TTR), which is the ratio of unique word forms (*types*) to the number of all running words in a text (*tokens*). When comparing texts of different lengths, the ratio must be normalized in some way, as the results would not be truly comparable (as corpora get larger, the number of tokens can become huge, but the number of types cannot follow the same pace, i.e. it does not rise proportionally to the number of tokens). In this case, for normalization purposes, it is possible to use the standardized TTR (first introduced by Johnson (1944); implemented in *WordSmith Tools 4.0* by Scott (2004)). When applying the sTTR method, the corpus is first cut into parts of the same length, i.e. the same token count, and then the TTR is consecutively measured for the so-obtained chunks and standardized, i.e. averaged, at the end of the procedure. However, the cutting is mechanical and the so-obtained chunks may vary greatly vocabulary-wise and not be very representative of the entire corpus. This is why Covington and McFall (2010) suggested a modification of the method: changing the first stage of the procedure, i.e. the method of cutting the text. They propose averaging “moving” chunks: e.g. if we decide

that the size of the chunk is 1,000 words, then the first chunk would comprise the text from the 1st to the 1,000th word, the second one would contain the text found between the 2nd and the 1,001st word, the third one would feature the text between the 3rd and the 1,002nd word, and so on, until the whole of the corpus is exhausted. In this way, substantially more chunks are obtained than by just cutting the text into consecutive stretches of equal length. For instance, in a corpus of 100,000 words, by applying the STTR cutting method, we will obtain 100 chunks of 1,000 words. However, by applying the cutting method as proposed by Covington and McFall (2010), we would obtain 99,001 chunks of the same length, and averaging this many chunks allows for more precision. Covington and McFall (2010) believe that the measures obtained in this way are more representative of the entire corpus.

2.3. Academic vocabulary

Academic vocabulary refers to words commonly found in academic texts. It is also sometimes called sub-technical vocabulary (Cowan, 1974; Anderson, 1980; Yang, 1986; Baker, 1988). Some authors have identified such vocabulary outside the most frequent words of English and have put it in specialized word lists. The first influential one was the University Word List (UWL), compiled by Xue and Nation (1984) as a combination of its four predecessor lists (Campion & Elley, 1971; Praninskas, 1972; Lynn, 1973; Ghadessy, 1979). It was followed by the well-known Academic Word List (AWL), built by Coxhead in 2000, who first excluded the highly frequent general words from her corpus (as represented by the GSL, in this concrete case). The coverage of the AWL in various academic corpora is typically around 10% (Coxhead, 2000; Chen & Ge, 2007; Valipouri & Nassaji, 2013, etc.). Although the list has been somewhat criticized, e.g. for not being extensive enough (Hancioglu and Eldridge, 2007) and for not providing enough specificity as per disciplines (Hyland & Tse, 2007), it has been fairly influential in the world of lexical profiling, as well as in EAP teaching and learning, and many lists produced since have followed it as a role model. Such are the new academic word lists: one produced by Browne, Culligan and Phillips (NAWL, 2013), and the other called the Academic Vocabulary List (AVL), produced by Gardener and Davies (2013), both from much larger corpora than the original AWL. Since the AWL has been used in many previous studies, using this list allows for comparisons with the results obtained earlier, which is an advantage for our research purposes.

Some authors have produced discipline-specific academic word lists, whereas others have embarked on generating more general ones, covering several fields: among these lists is the Pilot Science List (Coxhead & Hirsh, 2007), covering almost 4% in the science corpus consisting of 14 disciplines. The list was built by first excluding the GSL and the AWL words.

Another list is the Academic Article Word List for Social Sciences (AAWL-SS), produced by Kwarý and Artha (2017). Like the AWL, it was built after the exclusion of the GSL words from a 1-million-word corpus. Although it contains just 350 types, the authors argue that it has a better coverage in their corpus than the AWL, which has considerably more items (12.65% vs. 11.76%). However, we determined that some of the types from the AAWL-SS could be added to the GSL, as missing family members (e.g. *additionally*, *competitive*, *determinant*, *developmental*, *organizational*, *reference*, *settings* and *tourism*), if we apply the commonly used rules and definitions devised by Bauer and Nation (1993), with the *level 6* definition of word family (this includes derivation using all most frequent, productive and regular prefixes and suffixes). These types and members derived from them belong to high-frequency vocabulary, which may be a significant factor when comparing the efficiency of word lists. In the case of the AWL and the GSL there is no overlapping (the AWL applies Bauer and Nation's definitions) and, therefore, on methodological grounds, we find that a direct comparison between the AAWL-SS and the AWL cannot be made before addressing the issue mentioned above.

3. Methodology, corpus and research questions

In this paper we intend to study the lexical profile of literary academic articles. For that purpose, we use a corpus containing 110 literary academic articles, taken from reputable 10 literary studies journals, all indexed in Clarivate Analytics (11 articles per journal) and published in the years 2018 and 2019. The articles were converted to plain text files and the reference sections were removed so as to reduce the load of proper nouns, which is a common practice in lexical profiling studies. The corpus obtained in this way contains 822,859 tokens.

For comparison purposes, we use two additional corpora in this study.

The first is the OA STM Corpus (Open Access Corpus of Scientific, Technical, and Medical Content), freely-distributable from Elsevier Labs¹ (2015), which contains 11 articles from each of the following ten domains: agriculture,

astronomy, biology, chemistry, computer science, earth science, engineering, materials science, mathematics and medicine (a total of 754,593 tokens).

We compiled another corpus in a similar way, using 110 recently-published articles from 10 high-impact journals indexed in Clarivate Analytics. The journals are from the following disciplines: anthropology, archaeology, economics, history, law, linguistics, political science, psychology, philosophy and sociology (literary studies were deliberately excluded). We call this the Social Sciences and Humanities Research Articles Corpus (SSH-RAC), containing 952,181 running words.

All three corpora used in this study contain 110 articles which have been published recently in reputable journals. Reference sections, tables, figures, running titles etc., have all been removed from all the three corpora.

We use the lexical frequency profiling methodology, where we first analyse our corpora against the set of 25 word lists generated by Nation (2012) from the combined corpus of the BNC and the COCA. The aim is to see how much corpus coverage is reached with each of the individual lists and whether reading comprehension may be achieved using general high- and mid-frequency vocabulary solely. Specifically, we want to determine how many word lists, or thousands of word families, are needed to reach the coverages of 95% and 98%, respectively, needed for adequate and optimal reading comprehension.

We then calculate the standardized TTR ($sTTR$) and the moving average TTR ($maTTR$), as measures of lexical diversity of this genre, and see how they correspond to the vocabulary level data.

Finally, we determine the presence of academic vocabulary. We use Coxhead's AWL (2000) and the AAWL-SS (Kwary & Artha, 2017). These are available as lists of headwords (e.g. *abandon*, *abstract*...) and needed to be expanded into an all-member format (e.g. *abandon*, *abandoning*, *abandons*, *abandonment*; *abstract*, *abstracted*, *abstractedly*, *abstracting*, *abstraction*, *abstractions*...), for the calculation of their coverages in the corpora. This was done by means of Cobb's Familizer (2018).

The research questions posed in this study are the following:

1. How many words are needed to read literary academic articles?
2. How lexically diverse is the language of literary academic articles?

3. How much academic vocabulary is contained in literary academic articles?
4. How similar or different is the lexical profile of literary academic articles to research articles from the STM field, on the one hand, and social sciences and other humanities, on the other hand?

The software used for lexical profiling is the AntWordProfiler 1.4.0 (Anthony, 2014). For calculating the sTTR and maTTR, we first use AntFileSplitter (Anthony, 2017) and then MaWaTaTaRaD (Milička, 2013).

4. Results and analysis

Our analysis starts with Table 1, in which we present the results regarding the coverage of general-purpose vocabulary, as represented by the GSL, and the academic vocabulary, as represented by the AWL, in our three corpora.

Word lists	Literary studies	STM sciences	Social sciences & human.
GSL	74.37	66.73	73.46
AWL	6.92	10.24	10.91

Table 1. Coverage of GSL and AWL in the three corpora (%).

As shown in Table 1, the best coverage of the GSL was found in literary studies articles, followed by papers in social sciences and humanities, whereas in the STM field the coverage is substantially lower. Thus, literary studies articles feature a lexical profile which is similar to social sciences and other humanities, in the category of general-purpose vocabulary, with a slightly greater coverage of the GSL when compared to them.

The findings regarding the GSL's coverage are similar to those found in other studies. For instance, the ones Coxhead (2000) determined in her 3.5-million-word academic corpus, which consisted of four subsections: arts, commerce, law and science. The overall coverage of the GSL in her corpus was 76.1% and, for the subsections, the respective results were as follows: arts – 77.4%, commerce – 76.8%, law – 79.1%, and science – 72.7%. Her coverages were somewhat higher than the ones we obtained, which is explained by the differences in our corpora, namely, Coxhead used first-year university texts, as opposed to our research articles corpora, which have a

more complex vocabulary as they are intended to be read by graduates and professionals, rather than undergraduates. Still, we note the same tendencies in both studies: the coverage of general-purpose vocabulary seems to be higher in social sciences and humanities, as opposed to the STM fields. This is likely due to the fact that the STM fields tend to have a more specialized terminology, which accounts for a substantial share in their vocabulary coverage. For instance, Nation (2013: 31) finds that disciplines such as botany or medicine have large technical vocabularies, well exceeding 6,000 words, but that disciplines such as applied linguistics likely have a much smaller technical vocabulary, comprising between 1,000 and 2,000 words.

As for the AWL's coverage, it is significantly lower in literary studies (6.92%) than in the STM field, on the one hand, and social sciences and other humanities, on the other (10.24% and 10.91%, respectively), which is an important finding. This can be attributed to the specificity of the structure of literary papers whose meta-language is much less formulaic. The articles of social sciences, other humanities, and the STM field have a more rigid structure and repetitive academic vocabulary than is the case with literary studies articles, which are sometimes, on account of this, called simply 'essays'. The fact that research articles in literary studies display some genre peculiarities certainly plays a role in the choice of vocabulary. As a non-negligible influence in this type of scholarly articles one must also take into account the specific kind of corpora which impacts the researcher on both conscious and, possibly, subconscious levels. The language of *belles lettres*, which is the core research interest and which the scholars are exposed to (both in corpora and other research articles), undeniably influences the meta-language of these articles.

The level of academic vocabulary found in social sciences and other humanities, and the STM field, is in line with the results of other studies (Coxhead, 2000; Chen & Ge, 2007; Valipouri & Nassaji, 2013, etc.) and so is the finding that social sciences and other humanities use somewhat more academic vocabulary than do the STM disciplines (Kwary & Artha, 2017). The genre of literary studies articles and its vocabulary is indeed different in this respect and this could have significant pedagogical implications.

For illustration purpose of what kind of vocabulary is represented by the GSL and the AWL in our corpus, we present the extract below (taken from Sibanda (2018)). The words belonging to the GSL are underlined, whereas those belonging to the AWL are in bold.

It is my argument, however, that such an assertion is far too simplistic to explain the extensive use of negative stereotyping of the black self without providing a more nuanced engagement with the realities of postcolonial subjectivities. Bulawayo’s novel harkens back to the style of resistance writing that was predominant in South Africa during the era of apartheid, especially among black authors whose writing was so overtly political as to undermine its aesthetic appeal. Whilst it is not the suggestion of this article that Bulawayo’s novel is lacking in aesthetic appeal, I argue that it is the form of writing that propelled Njabulo Ndebele to write Rediscovery of the Ordinary (1986) and for Albie Sachs to present a paper at an ANC conference entitled Preparing Ourselves for Freedom: Culture and the ANC Constitutional Guidelines (1991). In both instances, they were engaging with the ways in which writing needs to move away from overt politicization and opt for a subtler aesthetic that is reflective of realities of people’s lives without relinquishing the role of the writer in society. But Bulawayo reflects the worst excesses of poverty and “defines [it] as merely the observable suffering resulting from a simple lack of material resources” (Roenigk, 2014: 1). The result of this approach is precisely what is termed poverty porn because she fails to reflect in her novel the act that “[p]overty is a result of both individual and systemic problems, involving not only personal circumstances but the social and justice systems in place that either works to empower the poor or perpetuate their conditions” (Roenigk, 2014: 1).

(An extract from Sibanda (2018), published in the Journal of Literary Studies)

As we can see, a great number of words are covered by the said two lists. Some of the remaining words have no learning load, such as some proper names (e.g. *Africa, Roenigk, Ndebele, Bulawayo*, etc.) and abbreviations (e.g. *ANC*). However, even the combined two lists are insufficient for the lower threshold of reading comprehension, bearing in mind that some words which are important for understanding the text are outside them (e.g. *postcolonial, subjectivity, novel, aesthetic, appeal, subtle, reality*).

Our analysis proceeds with Table 2, in which we present the coverage of the academic vocabulary typical of social sciences, as represented by the AAWL-SS word list (Kwary & Artha, 2017).

Word lists	Literary studies	STM sciences	Social sciences & human.
AAWL-SS	5.49	8.49	9.22

Table 2. Coverage of AAWL-SS in the three corpora (%).

As can be seen from Table 2, the coverage of AAWL-SS in the three corpora coincides with the AWL's results, which additionally corroborates the findings presented above regarding a strikingly lower presence of academic vocabulary in literary research articles as opposed to research articles of social sciences and other humanities, as well as the STM disciplines. We will not compare the results of STM disciplines and social sciences and humanities, as the word list used was extracted from social sciences and cannot be fairly applied to the STM field.

Table 3 presents the coverages of the word lists derived from the BNC/COCA corpus, together with the words which have no or have minimum learning load: these are proper nouns, which are typically easy to identify; abbreviations, which are generally explained the first time they are used in a text; and marginal words, which contain the letters of the alphabet, as well as swear words and exclamations (the latter two are not covered in our three corpora).

BNC/COCA word lists	Literary studies	STM sciences	Social sciences & human.
2,000 + proper n., abbrev. and marginal words	80.96	73.7	80.01
3,000 + proper n., abbrev. and marginal words	89.22	83.52	90.24
4,000 + proper n., abbrev. and marginal words	91.73	86.8	92.61
5,000 + proper n., abbrev. and marginal words	92.98	88.68	93.74
6,000 + proper n., abbrev. and marginal words	93.92	89.77	94.48
7,000 + proper n., abbrev. and marginal words	94.63	90.36	95.05
8,000 + proper n., abbrev. and marginal words	95.16	90.92	95.75
9,000 + proper n., abbrev. and marginal words	95.49	91.29	96.03
25,000 + proper n., abbrev. and marginal words	96.81	93.37	96.51

Table 3. Vocabulary coverage in three corpora (%).

Results coinciding with those presented earlier were obtained when the three corpora were compared against the BNC/COCA word lists: the lower reading threshold of 95%-coverage for literary articles is achieved at 8,000 + proper nouns, abbreviations and marginal words, whereas articles in the area of social sciences and humanities could be read at 7,000 + proper nouns, abbreviations and marginal words. However, not even the entire set of 25,000 + proper nouns, abbreviations and marginal words will suffice to read

the STM articles even barely comfortably, as they together cover just 93.37%. Our results differ from those achieved by Hsu (2011), who found that 5,000 words were needed to reach the coverage of 95% in economic research articles, but do coincide with Hsu (2013) and Vuković-Stamatović (2020).

Considering that a native speaker who has graduated from high school knows about 20,000 word families and that even illiterate native speakers are expected to know the first 9,000 word families of English (Nation, 2013: 26), this means that literary research articles can be adequately read by both the educated and the uneducated native speakers of English. As would be the case with all complex academic texts, this would probably entail some guesswork from the context and using a dictionary, generally without outside assistance, especially in the second group. On the other hand, only very proficient ESL speakers have a vocabulary of 9,000 word families, which is rarely reached by non-natives, and these would typically be students of doctoral studies taught in English and professionals (Nation, 2013: 26). As many non-native adults will only have a vocabulary of less than 5,000 word families (Nation & Waring, 1997), this means that reading literary research articles (as well as articles from the remaining social sciences, humanities and the STM disciplines) will be beyond their reach and will typically be reserved for highly proficient graduates and professionals. Research articles are mainly read by these two target groups anyway, but a problem may arise if teachers ask undergraduate non-native speakers of English to use research articles for their seminar and graduation papers, as these would, in all likelihood, not be readable to the degree needed. Certainly, testing students' vocabulary level prior to assigning such tasks would be advisable (for instance, using the Vocabulary Levels Tests, initially devised by Nation (1983)).

The level of 98%, needed for ideal reading, remains, however, unreached by any group. In all probability, even educated native speakers will initially need some assistance when starting out to read these research articles. For non-native speakers, this means that most of them will need to use a dictionary or some other form of assistance, if they wish to fully cover the entire vocabulary used in those articles.

As expected, the lexical profile of literary research articles is more similar to that of social sciences and other humanities, than to the STM disciplines. Still, the vocabulary of literary academic articles seems to be somewhat more complex than that of social sciences and other humanities, in that it requires more words for the same level of reading comprehension. As noted earlier,

the language of literary academic papers is less formulaic and we would say that it often strives to emulate the language of the narrative genres it focuses on. The style of writing is very important in this field (some would say as important as its content), which is why repetitiveness is avoided. Its abundance of synonyms, figures of speech, such as metaphor, metonymy, allegory and other literary devices, undoubtedly influence its vocabulary complexity and raise its vocabulary level.

It would appear that the STM disciplines are the most demanding in terms of vocabulary level, as they do not reach the adequate reading comprehension threshold even with all the lists derived from the BNC/COCA corpus. However, we must understand that these lists represent words which are most frequent in general English, as represented by the two reference corpora they were derived from, and that the STM disciplines, beyond some level of general high-frequency vocabulary, use a lot of specialized vocabulary which is highly infrequent in general language (for instance, names of chemical compounds, minerals, alloys, etc.). So, taken only by itself, our finding that all BNC/COCA lists cover just 93.37% of the words used in the STM research articles cannot be taken as proof that the vocabulary of STM disciplines is somehow more complex than that of the other sciences and humanities.

To illustrate the coverage of the BNC/COCA lists in our corpus and what kind of words these lists represent, we will use another extract from the literary academic articles corpus (taken from Spadaro, 2019: 21). The figure was created by means of the software *AntWordProfiler*. The colours mark the following word lists: red – 1st 1,000 words, green – 2nd 1,000 words, blue – 3rd 1,000 words, pink – 4th 1,000 words, violet – 5th 1,000 words, orange – 6th 1,000 words, brown – 7th 1,000 words, yellow – 8th 1,000 words, dark blue – 9th 1,000, black – the remaining lists. Words in black were not listed in the said word lists.

To Taylor, Hoskyns' poem is in perpetual motion: it kicks, flings, winces and shifts. These oscillations stem from its lack of an ontological base, its lack of 'sense, or reason, forme, or hue.' Yet, according to Taylor's estimate, Hoskyns' incongruities find shape through Crudities: they manifest to form a portrait of 'thy book and thee.' Through highlighting the vacuity of their work, and identifying Coryat and Crudities as the reality against which it becomes legible, Hoskyns and Taylor accentuate absence. This absence, and its accompanying intra-textual relationality, is a microcosm for grotesque literatures, whose reliance on genre imbues them with a sense of hollow formlessness. Moreover, Taylor's poem betrays what Bernard Clapp describes as 'a lasting unease about his social and cultural identity,' torn between his identity as an 'unsophisticated waterman' and his growing court connections. Through identifying his work as an imitation of Hoskyns, Taylor classes himself as an 'ape', mimicking the actions of a superior. This humble socio-economic background has led Rebecca Fall to suggest that Taylor promoted 'a democratic model of poetics that prioritises access over understanding.' If such a model exists, however, it is not the invention of Taylor but instead an attribute of all grotesque forms: Taylor's 'impenetrable discourse' aligns with the broader relationality of grotesque literature, continually relocating meaning beyond the parameters of the text. So, in referencing himself as an ape imitating Hoskyns, Taylor attaches his poem to that of his predecessor and identifies it as the nucleus from which his own work has sprung. These layers of relationality – connecting Taylor, Hoskyns, Coryat and, of course, poetic form – render the grotesque text perpetually open, drawing a network of meaning that transcends itself. This is typical of Taylor's work more broadly, and in particular his linguistic inventions, many of which rely on, and allude to, the travels of Coryat. Overtly resisting comprehension, nonsense verse evokes longing for a site of decorum, of meaning and cohesion: that is, genre.

Figure 1. An extract from Spadaro (2019: 21).

In the first 1,000 words we find all the functional words (such as *of, the, and, that...*) and many other general words which pervade the text. In the 2nd 1,000 words, there are words such as *poem, lack, according, identify, lasting, unease, social*, etc. In the 3rd 1,000 words, we find words such as: *estimate, accompanying, absence, moreover, unsophisticated, superior*, etc. The words from the 4th 1,000 words are: *flings, thee, genre, betray, humble, parameter, predecessor, longing, imitation* and *nonsense*. The 5th 1,000 words are represented by: *mimicking, nucleus, overtly* and *transcend*. The 6th 1,000 words include: *winces, perpetual, perpetually, oscillation, hue, motion* and *ape*. The next 1,000 words, i.e. the 7th BNC/COCA list contains: *incongruities, grotesque* and *cohesion*. From the 8th 1,000 words, in this extract we find just the word *ontological*. In the 9th 1,000 words we come across: *impenetrable, imbues* and *legible*. The 10th BNC/COCA list is represented by just one word in this extract (*microcosm*), whereas from the 11th word list we find two: *poetics* and *decorum*. The remaining words are: *forme* (16th 1,000), *vacuity* (18th 1,000) and various proper names (Taylor, Hoskyns), which belong to the proper names list. Some words are not found in the BNC/COCA set, as these lists are still imperfect. For instance, two proper names, Clapp and Coryat, are not found in Nation's proper names list (this list can never be complete); 'sense' is an abbreviated form quoted so from Taylor, whereas *relationality* is a low-frequency member of the word family whose headword is *relate* and as such should be added to the 1st 1,000 words. As we said, 8,000 word families are needed to read literary journal articles adequately. This would mean that words such as *impenetrable, imbue, legible, microcosm, poetics, decorum, forme*, and *vacuity* could be guessed from the context or looked up in a dictionary by the reader, without outside assistance, or some of them could simply be skipped, without significantly endangering the reading

comprehension of the entire text. Still, as can be seen from this example, reading would be much more laborious than in the case the reader already knew most of these eight words.

What one can notice here based on this extract and what is generally true of all the three corpora used in this study is the vast presence of Latin- and Greek-based vocabulary, especially the former. Virtually all the words above from the 3rd frequency band onwards have Latin and Greek roots, with rare exceptions (the exceptions being just *moreover* (3rd band); *fling*, *thee*, *longing* (4th band); and *huc* (6th band)). Bird (1987) found that in the most frequent 1,000 words of English, the words of Italic and Hellenic origin made up 40% of the words; this percentage, however, rose in higher frequency bands, where their share averaged 60%. The share of Latin also depended on the field, Rocha-e-Silva (2018) found. He inspected medical, financial, sport, literary and colloquial corpora, and found that most Latin-based words were used in the medical and financial texts, while they were the least frequent in the literary and colloquial corpora. Green (2015) estimates that 90% of the vocabulary of science and technology has Latin and Greek roots. This fact certainly represents an advantage for those learners of English who speak languages based on Latin, i.e. Romance languages. For instance, Moss (1992) finds that in scientific and technical texts, Spanish speakers can expect 30% of cognates, which they can successfully recognize in 60% of the cases. Quero (2015) studied English medical texts and found that 55% of the words used were English/Spanish cognates. Moreover, the share of Graeco-Latin in English, especially in science, also has important pedagogical implications for the speakers of those languages that have heavily borrowed from Latin. Rocha-e-Silva (2018: 2) comments that all European languages have borrowed immensely from Latin, as a consequence of the Renaissance, and the industrial and scientific revolutions. We may thus speak of the so-called *International Scientific Vocabulary*,² a set of mostly Latin- and Greek-based scientific-technical words, typically phonologically and morphologically adapted to various languages, but used with the same meaning across them. As we have already suggested, testing the learners before assigning any task of reading research articles is strongly advisable; depending on the other languages they speak, there could be a substantial number of cognates and borrowed words they could successfully recognize in context and this could reduce the figures regarding the vocabulary load. Some authors have even excluded Latin words when making a word list for their students; for instance, Konstantakis (2007) does so when he sets out to

derive his business word list, arguing that these words have a low learning burden. However, to be able to do this, one needs to have a particular learning group in mind, typically one that is homogenous in terms of their linguistic background.

Our final table, Table 4, presents the results regarding the lexical diversity as displayed in the three corpora. As commented in the Methodology section, for these purposes we use both the standardized and the moving average type-to-token ratio (calculated for chunks comprising 50,000 tokens).

Word lists	Literary studies	STM sciences	Social sciences & human.
sTTR	0.152	0.095	0.119
maTTR	0.154	0.097	0.122

Table 4. Lexical diversity in the three corpora (%).

Table 4 shows the lexical diversity of the three corpora. As was mentioned in the Methodology section, vocabulary diversity does not necessarily coincide with vocabulary level and this is clearly shown in our study. Although literary papers could be read at 8,000 word families, and papers in social sciences and other humanities at 7,000 word families, these papers exhibit significantly higher vocabulary diversity than the STM papers, which cannot be read even at 25,000 most frequent word families in English. What our results indicate, when considered jointly, is not that the STM papers are necessarily more demanding in terms of vocabulary, i.e. that more words are needed to read them. It is a more specific vocabulary that is needed to read them successfully or comfortably, which further stresses the need for creating specific word lists for the STM disciplines, bearing in mind that the specific vocabulary is repetitive in them (as confirmed by the sTTR/maTTR results).

Nevertheless, when the results on both vocabulary level and vocabulary diversity are considered jointly, literary studies articles could be considered the most lexically demanding articles.

5. Pedagogical implications

The results presented in this study could be of particular interest to non-native learners of English literature studies and their teachers, as an

indication of whether and to what extent learners can cope with reading literary academic articles or when they will be able to do so. It would be best to first test these learners' vocabulary using, perhaps, Nation's vocabulary tests (2013). Depending on other languages they speak, they could be able to recognize a number of cognates and loanwords (especially from Latin and Greek), which means they would face a lower vocabulary load than might be suggested by the results reported here. By knowing how many thousands of words learners know, educators will be able to tell in advance if students can adequately and optimally read these texts or if they are too much of a challenge considering their current vocabulary knowledge. These data will help teachers decide if and when they can make use of academic articles as learning materials and at what level. It is highly likely that only post-graduates will be able to read and benefit from scholarly articles in literary studies.

The comparatively low level of the AWL in literary academic articles might mean that teaching it in this discipline would have a limited value since, as we have seen, it is used significantly less than in other two areas.

6. Limitations of the study

As stated in the introduction, nowadays there are more recent and up-to-date versions of general service and academic vocabulary lists than the ones we used (the GSL and the AWL); however, opting for the older lists in this case allowed for a comparison of our results with those from the previous studies.

As explained earlier, in this paper we used three corpora, each of them containing 110 articles; in the compilation of our corpora of literary articles and social science and humanities, we decided to apply the same make-up as the one employed in the OA STM Corpus (Elsevier Labs, 2015), in order to make the three corpora fully comparable. As a result, the three corpora are relatively large, each of them featuring between 750,000 and 1 million words. While they are all of a size which is sufficient for lexical profiling, they could also have been larger, taking into account that two of them represent a number of disciplines.

7. Conclusion

As shown in the results section, a corpus of 110 literary academic articles corpus (compiled by the authors) was subjected to lexical frequency profiling. The so-obtained lexical profile was subsequently compared to the lexical profile of two additional corpora, containing 110 research articles each: the STM corpus (Elsevier Labs, 2015) and the corpus of social sciences and humanities (also compiled by the authors). We analysed their vocabulary level, variation and how much academic vocabulary they featured.

We found that literary studies articles boast the largest coverage of general words, followed closely by research articles in social sciences and other humanities, whereas in the STM field their coverage is substantially lower. However, the results significantly differ for academic vocabulary, found to be rather lower in literary scholarly articles than in other disciplines. The wide discrepancy in the level of academic language in literary study articles, not only when compared to the STM disciplines, but also to the other disciplines in the field of social sciences and humanities is, in all probability, due to the fact that the language in these papers is much less formulaic and that it, to some extent, resembles the subject of study (different narrative genres); the style of writing and avoiding repetitiveness are esteemed very highly in this discipline. In addition, literary studies follow the genre norms of research articles somewhat more loosely than other disciplines.

The literary academic papers can be read at 8,000 word families, needed for “reasonable” comprehension (a coverage of 95%), while the articles from social sciences and other humanities can be adequately read at 7,000 word families. However, not even the whole set of the word lists derived from the BNC/COCA corpus was enough to read the STM research articles comfortably. The reason for this is that the STM field requires quite a specialized vocabulary which is situated beyond the highest frequency bands, but which is very repetitive, judging by its low score for lexical variation.

The combined results for the vocabulary level and vocabulary diversity suggest that literary studies articles are in fact the most lexically demanding articles amongst the ones we studied. Taking into account Nation’s (2013) findings on the number of word families normally known by educated L2 speakers, it is not until the post-graduate level of literary studies that educators may successfully use such articles with their students. There could be exceptions for students who are familiar with Latin and Greek roots, who could be able to successfully recognize some of the cognates and loanwords,

which would reduce their vocabulary load. This is why it is advisable to conduct a preliminary check of students' vocabulary level in order to decide if scholarly articles from a particular field can be used as their reading materials.

Based on our results, we believe that producing a discipline-specific word list for literary studies would be of much less use than for the STM disciplines, as there is much less repetitive vocabulary in them. Such a list, in all probability, would be much longer, i.e. contain many more words than those made for other disciplines, in order to reach some adequate level of coverage which would justify its making.

Article history:

Received 4 January 2020

Received in revised form 25 August 2020

Accepted 10 October 2020

References

- Afros, E. & C. F. Schryer (2009). "Promotional (meta)discourse in research articles in language and literary studies". *English for Specific Purposes* 28(1): 58-68. doi: <<https://doi.org/10.1016/j.esp.2008.09.001>>.
- Anderson, J. I. (1980). "The lexical difficulties of English medical discourse for Egyptian students". *English for Specific Purposes* 37(4): 3-5.
- Anthony, L. (2014). *AntWordProfiler* (Version 1.4.1) [Computer Software]. Waseda University. URL: <www.laurenceanthony.net/software>.
- Anthony, L. (2017). *AntFileSplitter* (Version 1.0.0) [Computer Software]. Waseda University. URL: <www.laurenceanthony.net/software/antfilesplitter/>.
- Baker, M. (1988). "Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles". *Reading in a Foreign Language* 4(2): 91-105.
- Balocco, A. E. (2000). "Who's afraid of literature? Rhetorical routines in literary research articles". *Especialist* 21(2): 207-223.
- Bauer, L. & P. Nation (1993). "Word families". *International Journal of Lexicography* 6(4): 253-279. doi: <<https://doi.org/10.1093/ijl/6.4.253>>.
- Bird, N. (1987). "Words, lemmas and frequency lists: Old problems and new challenges". *Al-manakh* 6(2): 42-50.
- Brezina, V. & D. Gablasova (2013). "Is there a core general vocabulary? Introducing the new general service list". *Applied Linguistics* 36(1): 1-22. doi: <<https://doi.org/10.1093/applin/amt018>>.
- Browne, C., B. Culligan & J. Phillips (2013). *The New Academic World List*. URL: <www.newgeneral.servicelist.org/nawl-new-academic-word-list/>.
- Campion, M. E. & W.B. Elley (1971). *An Academic Vocabulary List*. Council of Educational Research.
- Chen, Q. & G. Ge (2007). "A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles". *English for Specific Purposes* 26: 502-514. doi: <<https://doi.org/10.1016/j.esp.2007.04.003>>.
- Cobb, T. (2018). *Familizer + Lemmatizer v.2.0*. [Computer Software]. URL: <<https://www.lexutor.ca/>>.
- Covington, M. A. & J. D. McFall (2010). "Cutting the Gordian knot: The moving-average type-token ratio (MATTR)". *Journal of Quantitative Linguistics* 17(2): 94-100. doi: <<https://doi.org/10.1080/09296171003643098>>.
- Cowan, J. R. (1974). "Lexical and syntactic research for the design of EFL reading materials". *TESOL Quarterly* 8(4): 389-400.
- Coxhead, A. (2000). "A new academic word list". *TESOL* 34(2): 213-238. doi: <<https://doi.org/10.2307/3587951>>.
- Coxhead, A. & D. Hirsh (2007). "A pilot science-specific word list". *Revue Française de Linguistique Appliquée* 12(2): 65-78. doi: <<https://>>

doi.org/10.3917/rfla.122.0065>.

Coxhead, A. (2018). *Vocabulary and English for Specific Purposes Research: Quantitative and Qualitative Perspectives*. Routledge.

Francis, W. N. & H. Kučera (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin.

Gardner, D. & M. Davies (2014). "A new academic vocabulary list". *Applied Linguistics* 35(3): 305-327. DOI: <<https://doi.org/10.1093/applin/amt015>>.

Ghadessy, M. (1979). "Frequency counts, word lists, and material preparation: a new approach". *English Teaching Forum* 17(1): 24-27. DOI: <<https://doi.org/10.1017/9781316155752.006>>.

Green, T. M. (2015). *The Greek & Latin Roots of English* (5th ed.). Rowman & Littlefield Publishers.

Hancioğlu, N. & J. Eldridge (2007). "Texts and frequency lists: some implications for practising teachers". *ELT Journal* 61(4): 330-340. DOI: <<https://doi.org/10.1093/elt/ccm051>>.

Hsu, W. (2011). "The vocabulary thresholds of business textbooks and business research articles for EFL learners". *English for Specific Purposes* 30(4): 247-257. DOI: <<https://doi.org/10.1016/j.esp.2011.04.005>>.

Hyland, K. & P. Tse (2007). "Is there an academic vocabulary?". *TESOL Quarterly* 41(2): 235-253. DOI: <<https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>>.

Jacoby, S. (1987). "References to other researchers in literary research articles". *ELR Journal* 33(1): 33-78.

Johnson, W. (1944). "Studies in language behavior. I: A program of research". *Psychological Monographs* 56(2): 1-15.

Konstantakis, N. (2007). "Creating a business word list for teaching business English". *ELIA* 7: 79-102.

Kwary, D. A. & A. F. Artha (2017). "The academic article word list for social sciences". *MEXTESOL* 41(4): 1-11.

Lake, W. M. & V. Cortes (2020). "Lexical bundles as reflections of disciplinary norms in Spanish and English literary criticism, history, and psychology research" in U. Romer, V. Cortes & E. Friginal (eds.), *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*, 183-204. John Benjamins.

Laufer, B. (1989). "What percentage of text-lexis is essential for comprehension?", in C. Lauren & M. Nordman (eds.), *Special Language: From*

Humans Thinking to Thinking Machines, 316-323. Multilingual Matters.

Lewin, B. A. & H. Perpignan (2012). "Recruiting the reader in literary criticism". *Text & Talk* 32(6): 751-772. DOI: <<https://doi.org/10.1515/text-2012-0035>>.

Lynn, R.W. (1973). "Preparing word lists: a suggested method". *RELC Journal* 4(1): 25-32. DOI: <<https://doi.org/10.1177/003368827300400103>>.

Malášková, M. (2012). "Hedges as writer protective devices in applied linguistics and literary criticism research articles". *Discourse and Interaction* 5(1): 31-47. DOI: <<https://doi.org/10.5817/DI2012-1-31>>.

Nation, I. S. P. (1983). "Testing and teaching vocabulary". *Guidelines* 5: 12-25.

Nation, I. S. P. (2006). "How large a vocabulary is needed for reading and listening?". *Canadian Modern Language Review* 63: 59-82.

Nation, I. S. P. (2012). *The BNC/COCA Word Family Lists*. URL: <www.victoria.ac.nz/lals/about/staff/paul-nation>.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press.

Nation, I. S. P. (2016). *Making and Using Word Lists for Language Learning and Testing*. John Benjamins.

Nation, I. S. P. & R. Waring (1997). "Vocabulary size, text coverage and word lists". *Vocabulary: Description, Acquisition and Pedagogy* 14: 6-19.

McQuillan, J. (2016). "What can readers read after graded readers?". *Reading in a Foreign Language* 28(1): 63-78.

Milička, J. (2013). *MaWaTaTaRaD 2* [Computer Software]. Charles University & Olomouc: Palacký University. URL: <www.milicka.cz/en/software.htm>.

Moss, G. (1992). "Cognate recognition: Its importance in the teaching of ESP reading courses to Spanish speakers". *English for Specific Purposes* 11(2): 141-158. DOI: <[https://doi.org/10.1016/S0889-4906\(05\)80005-5](https://doi.org/10.1016/S0889-4906(05)80005-5)>.

Praninskas, J. (1972). *American University Word List*. Longman.

Quero, B. (2015). *Estimating the Vocabulary Size of L1 Spanish ESP Learners and the Vocabulary Load of Medical Textbooks*. Doctoral dissertation, Victoria University of Wellington.

Rocha-e-Silva, M. (2018). "Writing good English: is scientific English a Latin language in disguise?".

MedicalExpress 5: 1-5.

Schmitt, N. & D. Schmitt (2014). "A reassessment of frequency and vocabulary size in L2 vocabulary teaching". *Language Teaching* 47(4): 484-503. DOI: <<https://doi.org/10.1017/S0261444812000018>>.

Scott, M. (2004). *Oxford WordSmith Tools* (version 4.0). Oxford University Press.

Shaw, P. (2009). "Linking adverbials in student and professional writing in literary studies: What makes writing mature" in M. Charles & D. Pecorari (eds.), *Academic Writing: At the Interface of Corpus and Discourse*, 215-235. Continuum.

Sibanda, S. (2018). "Ways of reading blackness: Exploring stereotyped constructions of blackness in NoViolet Bulawayo's *We Need New Names*". *Journal of Literary Studies* 34(3): 74-89. DOI: <<https://doi.org/10.1080/02564718.2018.1507155>>.

Spadar, K. L. (2019). "On classification and the grotesque: Theorizing para-genre in early modern nonsense verse and Montaigne's *Essais*". *Journal of Language, Literature and Culture* 66(1): 16-30. DOI: <<https://doi.org/10.1080/20512856.2019.1595480>>.

Tankó, G. (2017). "Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations". *Journal of English for Academic Purposes* 27: 42-55. DOI: <<https://doi.org/10.17507/jltr.1205.21>>.

Valipouri, L. & H. Nassaji (2013). "A corpus-based study of academic vocabulary in chemistry research articles". *Journal of English for Academic Purposes* 12(4): 248-263. DOI: <<https://doi.org/10.1016/j.jeap.2013.07.001>>.

Vuković-Stamatović, M. (2020). "Vocabulary complexity and reading and listening comprehension of various physics genres". *Corpus Linguistics and Linguistic Theory* 16(3): 487-514. DOI: <<https://doi.org/10.1515/cllt-2019-0022>>.

West, M. (1953). *A General Service List of English Words*. Longman.

Wilder, L. (2012). *Rhetorical Strategies and Genre Conventions in Literary Studies: Teaching and Writing in the Disciplines*. Southern Illinois University Press.

Wilder, L. (2005). "The rhetoric of literary criticism' revisited: Mistaken critics, complex contexts, and social justice". *Written Communication* 22(1): 76-119. DOI: <<https://doi.org/10.1177/0741088304272751>>.

Xue, G. & I. S. P. Nation (1984). "A university word list". *Language Learning and Communication* 3(2): 215-229.

Yang, H. (1986). "A new technique for identifying scientific/technical terms and describing science texts". *Literary and Linguistic Computing* 1(2): 93-103. DOI: <<https://doi.org/10.1093/lc/1.2.93>>.

Vesna Bratić is an Associate Professor of English Literature at the Faculty of Philology, University of Montenegro. She teaches undergraduate courses in 18th and 19th century English literature and graduate courses in American Literature at the English Department and ESP at the Faculty of Electrical Engineering. Her professional interests lie within contemporary American and English literature, gender studies, metaphor and discourse analysis.

Milica Vuković Stamatović, PhD, is an Associate Professor of English Language and Linguistics at the Faculty of Philology, University of Montenegro. She currently teaches Discourse Analysis and English for Specific Purposes. She has taught English for Physics and English for Mathematics for over a decade. Her main research interests lie in the field of discourse analysis, pragmatics and ESP.

NOTES

¹ URL: <<https://elsevierlabs.github.io/OA-STM-Corpus/>>.

² The term was first introduced by Gove in *Webster's Third New International Dictionary* (1961).