Vocabulary profile, lexical density and speech rate in science podcasts: How appropriate are science podcasts for EAP and EST listening?

Milica Vuković-Stamatović & Dragana Čarapić
University of Montenegro
vmilica@ucg.ac.me & draganac@ucg.ac.me

Abstract

One potential source of authentic input for listening activities in English for Academic Purposes (EAP) and English for Science and Technology (EST) could be science podcasts. In this study, we examine the vocabulary level of this genre, as well as the amount of specialized academic and scientific vocabulary it contains. Additionally, we study its lexical density and speech rate. These properties are examined in a corpus of 126 science podcasts taken from the journal *Nature*, the *American Association for the Advancement of Science (AAAS)* and the magazine *Scientific American*. These are compared with a mixed-disciplinary corpus of lectures (the MICASE lectures) and a corpus of science lectures (the MICASE's science-lecture subsection). On the one hand, science podcasts were found to contain a substantial amount of both academic and scientific vocabulary, and to have a very similar lexical density to that of lectures. On the other hand, their vocabulary level and speech rate were somewhat higher than those of lectures. Based on the analysis, we recommend the use of science podcasts for both EAP and EST listening activities for advanced learners. For learners with lower proficiencies, vocabulary support and/or adjusting the speed of the science podcasts is advised.

Keywords: science podcasts, lexical profile, lexical density, speech rate, academic listening.

Resumen

Perfil de vocabulario, densidad léxica y velocidad del habla en pódcast científicos: ¿En qué medida son apropiados los pódcast científicos para la comprensión oral en inglés con fines académicos (IFE) y en inglés científico-técnico (ICT)?

Los pódcast científicos pueden servir como fuente de obtención de *input* auténtico para actividades de escucha en inglés con fines académicos (IFE) y en inglés científico-técnico (ICT). En este estudio examinamos el nivel de vocabulario en este género, así como la cantidad de vocabulario académico y científico que contiene. Asimismo, estudiamos su densidad léxica y velocidad de habla. Estos parámetros se han examinado en un corpus de 126 pódcast científicos tomados de la revista *Nature*, de la *Asociación Estadounidense para el Avance de la Ciencia* y de la revista *Scientific American*. Este corpus se ha comparado con un corpus mixto de conferencias (las conferencias de MICASE) y otro corpus de conferencias científicas (la subsección de conferencias científicas de MICASE). Por un lado, se ha descubierto que los pódcast científicos contienen una cantidad considerable de vocabulario tanto académico como científico y presentan una densidad léxica muy similar a la de las conferencias. Por otro lado, su nivel de vocabulario y velocidad del habla es algo más alto que los de las conferencias. Con base en el análisis realizado, recomendamos el uso de pódcast científicos para actividades de escucha en IFE y ICT con estudiantes de niveles avanzados. Para alumnos con niveles de competencia más bajos se recomienda incorporar un apoyo para el vocabulario y/o ajustar la velocidad de los pódcast científicos.

Palabras clave: pódcast científicos, perfil léxico, densidad léxica, velocidad del habla, escucha académica.

1. Introduction

In teaching English for Academic Purposes (EAP) and English for Science and Technology (EST), the use of authentic teaching and learning materials is often recommended (Martínez, 2011; Benavent & Peñamaria, 2011). EAP and EST instructors and materials developers frequently find themselves in the challenging situation of having to find and choose appropriate authentic materials, including those intended for classroom listening and those advised for extensive listening at home (cf. Parkinson & Adendorff, 2004; Parkinson et al., 2007; Ye, 2020).

When selecting authentic materials for EAP and EST, a number of criteria may be taken into account. Among other things, the materials selected should be accessible to target learners in terms of their level of vocabulary (Nation, 2013; Coxhead, 2018) and they should also include some more specialised vocabulary, such as academic or technical vocabulary (cf. Coxhead, 2000; Nation, 2001; Coxhead & Hirsh, 2007). Other selection criteria may include the rate of delivery and lexical density (Wingrove, 2017; Liu, 2021), given that speech which is too fast or which includes too much information within

a certain timeframe may be difficult for some learners to understand and follow. Another important criterion would be that the materials should be engaging, to help learners' motivation. With many potential criteria to consider, the selection process can be difficult and time-consuming (cf. López-Carril et al., 2020).

Science and technology were an early focus of interest for English for Specific Purposes (ESP) researchers, which led to the expansion of EST and research in this field (Jenkins, 2012). In an effort to enhance EAP and EST instruction, applied linguists have shown an interest in studying the potential of using popular scientific genres as input (Myers, 2003; Parkinson & Adendorff, 2004; Parkinson et al., 2007). In this paper, we employ corpus-linguistic methods to study the suitability of science podcasts for listening purposes in EAP and EST. The EAP and EST courses which we here have in mind are those taught to university students, i.e. adults, who learn English as their L2. The study fits into a recent and expanding body of related studies, which includes: Csomay and Petrović's (2012) investigation into the legal vocabulary in movies and TV series on the topic of law and order, so as to assess these two genres' suitability for Legal English instruction, Wingrove's (2017) investigation into the use of TED talks for academic listening practice, Rolls and Rodgers' (2017) investigation into the suitability of science fiction and fantasy for EST, Hiltunen and Tyrkkö's study of academic vocabulary in Wikipedia articles for EAP purposes (2018), Vuković-Stamatović's study of the vocabulary in various physics genres (2020), Dang's analysis of the presence of specialised vocabulary in medical TV programmes (2020), Liu's investigation into the use of TED talks in English for Medicine (2023), Zhang's study of the suitability of TED Talks for English for Humanities (2022), and Vuković-Stamatović's study on the adequacy of science documentaries for EST (2022).

Podcasts are typically available as digital audio files which can be downloaded from the internet or streamed, usually in the form of a series, whose new episodes can be received automatically by their subscribers. The potential for using them to teach English has already been recognised – they have been found to improve the listening, speaking, pronunciation and grammar skills of English language learners (Lord, 2008; Putman & Kingsley, 2009; Knight, 2010; Fadda & Qasim, 2013; Li, 2012). In addition, many perception studies have found this genre to be appealing to learners and to increase motivation (Lee & Chan, 2007; Dlott, 2007; O'Bannon et al., 2011).

Science podcasts cover topics in physics, mathematics, chemistry, bioscience, medicine, pharmacology, geology, climate change, palaeontology, computer science, and engineering (MacKenzie, 2019). A vast majority of them are audio-only (87%) and U.S. productions (57%), according to MacKenzie's survey (2019). The same study also found that they are chiefly targeted at public audiences and typically hosted by scientists. The number of science podcasts has risen exponentially over the recent years.

The potential of using science podcasts for EAP listening has been indicated by Ye (2021), who performed a genre analysis of 110 episodes of the *Scientific American*'s podcast *60-Second Science*. As can be seen, this study focused on one particular science podcast series only, one which is not representative of the genre in terms of its duration – its short duration could, perhaps, result in an increase in its information load or speech rate. Moreover, the study did not use the type of corpus-linguistic methods applied in the studies listed above, which were entirely focused on determining the suitability of certain genres for some ESP areas. Therefore, there remains room for an additional investigation to complement Ye's study (2021) and give a more definitive answer to whether science podcasts are a good source of authentic material for EAP listening. The suitability of science podcasts for EST listening has been even less researched. Although some results point to their potential adequacy for EST listening (Chaikovska et al., 2019; Hawke, 2010), much more systematic research is needed to confirm this.

Bearing the above in mind, the aim of this paper is to contribute to answering how suitable science podcasts may be for EAP and EST listening in terms of their lexical profile, lexical density and speech rate. Inspired by Nesi (2001), the use of these three criteria for assessing a genre's suitability for specialised contexts of teaching English was applied by Wingrove (2017), and following Wingrove, by Liu (2021) and Vuković-Stamatović (2022). The choice of the three criteria as those which critically impact listening comprehension is further supported by other studies. Namely, Révész and Brunfaut (2013) and Brunfaut and Révész (2015) argue that lexical complexity and lexical density have the most adverse effect among the factors which may negatively impact language learners' understanding of audio and printed texts. Both of these aspects are explored in this study, with lexical complexity being investigated as part of determining the vocabulary profile. Zhao (1997) further identifies speech rate as a major factor affecting the listening comprehension of language learners.

In contrast to previous studies, this paper calculates lexical density for entire corpora and speech rate for complete science podcasts and lectures, rather than relying on 1-minute or 5-minute samples.

The research questions here asked, based on the aim stated above, are as follows:

1. How much vocabulary is needed for adequate listening comprehension in science podcasts compared to academic lectures in general and science lectures in particular?
2. How much-specialised vocabulary do science podcasts have compared to academic lectures in general and science lectures in particular?
3. What is the lexical density of science podcasts compared to academic lectures in general and science lectures in particular?
4. How similar is the speech rate of science podcasts to that of academic lectures?

2. Background

In this part of the paper, we present lexical profiling, lexical density and speech rate. We also present the genre of podcasts and how they may be used in ELT in general, and in EAP and EST in particular.

2.1. Lexical profiling, listening-comprehension vocabulary thresholds and word lists

Lexical profiling, a corpus-linguistic method developed by Laufer and Nation (1995), can be used to determine the lexical complexity of texts or how specialised they are vocabulary-wise. On the one hand, lexical complexity has generally been pointed out as one of the most powerful predictors of both reading and listening comprehension (cf. Nagy, 1988; Révész & Brunfaut, 2013; Brunfaut & Révész, 2015), with a linear relationship between vocabulary knowledge and level of comprehension (Schmitt et al., 2011; van Zeeland & Schmitt, 2013). On the other hand, in EST as an ESP area, a major focus is on learning specialised vocabulary and the materials used should generally reflect this fact (Nation, 2013).

The method of lexical profiling involves determining the vocabulary distribution in a given corpus based on its frequency in a reference corpus.

The texts which are lexically profiled are entered into a lexical-profiling software, along with various word lists – typically, sets of words belonging to various frequency bands (as determined in large reference corpora) or various specialised lists (e.g. academic vocabulary lists, discipline-specific word lists, etc.). The programme then determines the lexical coverage of the selected word lists, to create the lexical profile of the text analysed.

As one of the most well-known frequency-based vocabulary metrics, the lexical frequency profiling method is frequently used to assess how lexically demanding particular texts are. This method, although not the only one which can be used to determine lexical complexity, has been found to yield outcomes that are comparable to those of other methods (Lindqvist et al., 2013). The ability to compare a certain word-list's coverage in the text analysed against the reading and listening vocabulary thresholds, i.e. the minimum vocabulary coverages required for certain levels of comprehension (as reported in the literature), is one of the benefits of employing this method.

A vocabulary threshold for reading or listening comprehension refers to the minimum amount of vocabulary that a learner must know in order to read or listen to a specific text independently. Laufer (1989) argues that for a text to be reasonably understood, readers need to be familiar with at least 95% of the words in it, while Nation (2006) contends that for a text to be optimally understood, readers should be familiar with at least 98% of its words. In a series of experiments to determine the listening-comprehension vocabulary thresholds, van Zeeland and Schmitt (2013) concluded that 95%-vocabulary coverage is required for most listeners to achieve adequate comprehension, i.e. "good, but not necessarily complete comprehension" (van Zeeland & Schmitt, 2013, p. 475). They also found that many experiment participants could successfully listen to informal narratives with a 90%-vocabulary coverage.

We can determine the number of words required to reach these various levels of reading or listening comprehension in a particular text or speech, by comparing the lexical-profiling results against the vocabulary thresholds cited above. Under this method, words are often specified as either word families or lemmas. For example, *conclude, concludes, concluded, concluding, conclusion, conclusions* would all make one-word family (the "level-6 word family," as categorised by Bauer and Nation (1993)), while the lemma approach would see this set as having two lemmas, one of them

encompassing the inflected forms of the headword *conclude* and the other of the headword *conclusion*. The word-family approach is generally employed to determine the vocabulary load or level of texts (Coxhead, 2018).

There are various general-purpose word lists which can be used to determine the vocabulary level of texts, i.e. how many words are required to meet the various reading or listening comprehension thresholds in them. For lexical profiling, the most applicable is Nation's set of frequency-based word lists (2012). His 25-word lists in this set represent the top 25,000-word families in the BNC and the COCA. Nation (2012) also added four more lists to accompany the 25-word-list set – these contain abbreviations, proper names, marginal words (exclamations, swear words and letters of the alphabet), and compounds without hyphens.

There are also word lists with a narrower focus. The Academic Word List (AWL), created by Coxhead (2000), the New AWL (Browne et al., 2014), the Academic Vocabulary List (AVL), created by Gardner and Davies (2014), the Academic Spoken Word List (ASWL), created by Dang et al. (2017), are some of the word lists that contain the most common academic vocabulary. The most recent of them, Dang et al.'s ASWL (2017), represents spoken English as it is used in academic contexts. It was created using a corpus of 13 million words from 24 different academic disciplines. It was created "from scratch," which means that no word group was excluded in its making (unlike the AWL and the New AWL, which were derived from the words outside the GSL and New GSL, respectively). There are 1,741 word families in the ASWL, which cover 90.13% of the words in the corpus they originated from.

A number of word lists are based on even more specialised corpora, containing texts from one or several areas. The Science List (SL) by Coxhead and Hirsh (2007) is one of them. It does not contain the most frequent academic and general words. For our present paper, the most relevant is the Hard Science Spoken Word List (HSWL), developed by Dang (2018), from a 6.5 million token corpus of transcripts from 12 disciplines. This word list covers 90.94% of the words in the corpus from which it was created and it includes 1,595 word families that are the most common in hard science speech.

Specialised word lists may be used for selecting teaching and learning materials, given that their coverages in the texts analysed measure the presence of the most frequent specialised vocabulary.

2.2. Lexical density

One of the criteria for determining whether a certain genre is suitable for listening is calculating how lexically dense it is (Wingrove, 2017). Lexical density expresses the proportion of content words (nouns, verbs, adjectives and adverbs) to the total word-count in a text (Stubbs, 1986). It is a measure of how much information a listener must process in a certain amount of time (Wingrove, 2017).

The measures of lexical density are generally well-established for printed texts and some spoken genres (Mulyanti & Soeharto, 2020; Lukmana & Gunawan, 2021; Fadhil et al., 2023); however, there is very little research into the lexical density of the ESP listening materials or the materials considered for use in ESP. The present study is among the rare ones delving into this issue. The unique characteristics of listening make comprehending ESP listening materials a demanding task for language learners – as suggested earlier, of the factors that can negatively impact their understanding of audio and printed texts, lexical complexity and lexical density have been found to have the most adverse effect (Brunfaut & Révész, 2015; Révész & Brunfaut, 2013).

According to Wingrove's research (2017), the lexical density of the Yale-lecture series is 47.11%. Other findings from the literature, among others, are those by Stubbs (1986), who determined that the lexical density of non-fiction books ranged from 40% to 65%, while the lexical density of fiction works ranged from 40% to 54%; by Ure (1971), who found that speech typically has a lexical density below 40%; by Nesi (2001), who established that academic lectures feature a lexical density of 49%; and by Vuković-Stamatović (2022), who found that science documentaries feature a lexical density of 51.36%.

Research suggests that spoken texts generally have a lower lexical density than written texts (Ure, 1971; Halliday & Hasan, 2000; Johansson, 2008). There is, however, a lot of variation amongst different genres (Breeze, 2008) and this general assumption cannot be safely made for the genre of science podcasts, given that most of the talk in them is delivered by people from the academic and scientific world, on topics from specialised fields, laden with technical terms and abstract concepts. On account of this, we believe that investigating the lexical density of this particular genre is justified bearing in mind the aim of this study.

2.3. Speech rate

It is widely assumed that learners should be capable of grasping the meaning of the listening text at various speech rates (Abdurakhmonova & Abduvohidova, 2023). Speech rate, which has been identified as a major factor affecting the listening comprehension of language learners (Zhao, 1997), can be expressed in words per minute or syllables per second. It refers to the speed in which speech is delivered in relation to a unit of time (Tauroza & Allison, 1990). Griffiths (1990) considers that the speech rate of 100 w.p.m. may be considered slow, that of 150 w.p.m. may be considered average, while that of 200 w.p.m. may be considered fast. Tauroza and Allison (1990) provide more precise guidelines – slower than normal speech has a speech rate of less than 100 w.p.m., moderately slow speech ranges between 100 w.p.m. and 125 w.p.m., average speech rate would fall within the 125–160 w.p.m. span, moderately fast speech ranges between 160 w.p.m. and 185 w.p.m, while speech rates above the latter would be faster than normal.

Some speech-rate results from the literature for various genres include: the speech rate in the Yale-lecture series of 145 w.p.m. and in TED talks of 169 w.p.m. (Wingrove 2017); in radio broadcasts of 160 w.p.m., in conversation of 210 w.p.m., and in interviews of 190 w.p.m. (Tauroza & Allison, 1990).

It has been determined that, generally, slower speech rates contribute to better L2 listening comprehension (Griffiths, 1990, 1992). Likewise, some studies have proven that slowing down recordings or speech in the classroom can help learners improve their listening comprehension (Fujita, 2017; Griffiths, 1990, 1992; Hayati, 2010; Jensen & Vinther, 2003). While Griffiths' (1990, 1992) studies suggest a positive effect of slower speech, some researchers, such as Rader (1990), Fujita (2017) and Zhao (1997), emphasize that listeners' comprehension is additionally affected by other factors, among which lexical complexity and lexical density feature prominently, as suggested in the previous section. In addition, these factors themselves may be interconnected – namely, Nesi (2001) established a negative correlation between speech rate and lexical density, which means that faster speech tends to be less lexically dense and, *vice versa*, that slower speech tends to be more lexically dense. We will see whether this correlation holds for the genre of science podcasts and what pedagogical implications arise from the interconnection of the three factors affecting listening comprehension here explored.

2.4. Use of podcasts to teach English

Podcasts are available as either audios or videos (the so-called *vodcasts*), although they are primarily audio-only – this format is by far the most popular as it allows listening to podcasts while doing other activities (such as driving, walking, etc.) without having to look at a screen (MacKenzie, 2019). Due to this, we will focus on this format in our present review, as well as in our study. The term was coined in 2004, as a blend of *i-pod* and *broadcast* (Wu, 2008); the year of the birth of the term itself points to the fact that podcasts are a relatively recent phenomenon. Still, they have already managed to attract a lot of research attention.

What recommends the use of podcasts to teach English is the wide variety of content choice, portability and time-shifting opportunities (Fox, 2008). They may be used in class but also for extensive listening at home – individual listening to podcasts at one's own pace and at one's convenient time which has been found to be motivating in order to improve listening skills (Kavaliauskienė, 2008).

Podcasts may be used as a resource in teaching and learning English if they are selected according to the interest and level of learners (Rafique, 2019). As for interest, podcasts have been found to be appealing to learners and to increase motivation (Lee & Chan, 2007; Dlott, 2007; O'Bannon et al., 2011). As for the level, advanced learners can listen to authentic podcasts, while beginner and intermediate students may also be given some sections of the podcasts to listen to (Stanley, 2006). They have been found to improve the listening, speaking, pronunciation, vocabulary and grammar of the English language learners (Lord, 2008; Knight, 2010; Fadda & Qasim 2013; Li, 2012; Al-Ahdal, 2020, etc.). Podcasts may be used as supplements to textbook materials, authentic listening materials, a way for learners to obtain information on specific aspects of language use such as grammatical constructions or idioms, and also for speaking, in cases where students have to produce their own podcasts (Stanley, 2006).

In the EAP and ESP contexts, in particular, there are fewer studies on the use of podcasts. Ye (2021) performs a genre analysis of *Scientific American*'s *60-Second Science* podcasts. The author finds that science podcasts can be "placed towards the popular end on the professional-popular cline of scientific discourse" (2021, p. 10). The study determines that they frequently feature engagement devices such as inclusive pronouns, self-mentions, hedges and questions, and that they are highly dialogic, concluding that they are suitable

resources for EAP learners to acquire an awareness of academic spoken discourse. Bearing all this in mind, Ye (2021) concludes that science podcasts are adequate for EAP teaching and learning.

In EST contexts, Chaikovska et al. (2019) found that listening to podcasts on the topics from the field of electrical engineering improves electrical engineering students' listening comprehension and speaking. Wu (2008) finds that listening to science podcasts enhances cognitive engagement and higher-order thinking in students, which leads to improved teaching and learning of scientific English. Putman and Kingsley (2009) determined that teacher-produced science podcasts may enhance the acquisition of scientific vocabulary in EST. Hawke (2010) finds that science students can benefit from listening to science podcasts – however, his lower-proficiency learners needed supporting materials to aid the listening.

All these studies indicate that science podcasts should indeed be suitable for EAP and EST listening, but a more systematic study of this genre is needed in order to have a more informed judgement on the issue.

3. Data and method

We here provide the corpora details and the analytical procedure employed to answer the research questions posed.

3.1. Corpora

The corpora used in this paper are the following:

1. A corpus of 126 recent science podcasts collected for this study (most of them were published in 2022 and 2021). They range from 15 to 40 minutes in duration, but most are about 20 minutes long. The corpus was drawn from three sources:

- A total of 42 podcasts belong to the *Science Talk* series of the globally popular American science magazine – *Scientific American* (https://www.scientificamerican.com/podcast/science–talk/). The number of tokens in this part of the corpus is 159,233.
- 42 podcasts used in this corpus were published by the American Association for the Advancement of Science (AAAS), the largest general scientific society in the United States

(https://www.science.org/podcasts). The corpus contains the transcripts of the podcasts available on the stated websites. The number of tokens in this part of the corpus is 201,186.

- 42 science podcasts were taken from the website of the world-renowned journal Nature (https://www.nature.com/nature/articles?type=nature–podcast), the so-called *Nature Podcast* series. This corpus subsection includes 190,901 tokens.

The total number of tokens in our science-podcast corpus is 558,290.

MacKenzie (2019) demonstrates that the majority of the available science-podcast series produced in the English language are U.S. productions, which is why all three chosen podcasts series are American. Besides, these are certainly amongst the most popular science podcasts as they are produced by major and well-respected scientific organisations, magazines and journals.

2. All lectures available in the Michigan Corpus of Academic Spoken English (MICASE) corpus, containing 62 lectures with the token-count of 625,008 (https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple). The above corpus of science podcasts is compared against this mixed-disciplinary corpus of academic lectures to determine their suitability for EAP in general. The lectures come from a variety of disciplines and are sorted into four roughly equal sections: biological and health sciences, physical sciences and engineering, social sciences and education, humanities and arts. The MICASE lecture corpus was chosen as the reference lecture corpus, given that science podcasts in our corpus are American, so it made more sense to use the same English variety for our lecture corpus. Besides, its size is comparable to that of our podcast corpus.

3. The MICASE-corpus subsection of 31 lectures from biological and health sciences, and physical sciences and engineering (https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple), containing 289,131 tokens. The corpus of science podcasts is compared against this corpus of science lectures to determine their suitability for EST.

3.2. Data analysis

To answer research questions 1 and 2, we use the Lexical Frequency Profiling method, presented earlier (Laufer & Nation, 1995). As explained, the method involves loading a word list (or several lists) into a lexical-profiling

programme, as well as the corpus. The software provides the coverages of the loaded word list(s) in the corpus/corpora loaded. The lexical-profiling software which we have used here is *AntWordProfiler 1.4.1* (Anthony, 2014). The word lists used in this study are as follows (all of them were described in section 2):

- Nation's BNC/COCA word list set (2012)
- Dang et al.'s *Academic Spoken Word List* (ASWL) (2017)
- Dang's *Hard Science Spoken Word List* (HSWL) (2018)
- the *Science List* (SL) (Coxhead & Hirsh, 2007).

As for research question 1, the lexical coverages obtained for Nation's word list set (2012) are compared against two listening-comprehension thresholds relevant for listening comprehension (van Zeeland & Schmitt, 2013):

- the 90% coverage which ensures good comprehension for many listeners, and
- the 95% coverage which enables good comprehension for most listeners.

Research question 2 refers to specialised vocabulary. For this purpose, we use Dang et al.'s *Academic Spoken Word List* (ASWL) (2017), as this list was obtained from a corpus of spoken academic English, and Dang's *Hard Science Spoken Word List* (HSWL) (2018), as this list represents the most frequent vocabulary in spoken hard science. These two lists encompass high-frequency vocabulary (as they were built "from scratch", i.e. without the exclusion of any group of words) and so, to further determine the presence of the more specialised, i.e. technical vocabulary in science podcasts, we also used the SL (Coxhead & Hirsh, 2007). Despite the fact that the SL was derived from written academic-scientific language, its advantage here is that it was built outside the most frequent general-purpose and academic words (the GSL and the AWL), which is why it is more specialised, i.e. technical than the two lists above.

To answer research question 3, following Wingrove (2017), we use the online programme *Analyze my Writing* (http://www.analyzemywriting.com/index.html), which measures the percentage of content words in the total number of words in the corpus. While Wingrove (2017) and Liu (2021) used

samples, we used the entire corpus to calculate lexical density with the greatest certainty. Also, based on the individual results for every episode, we determine the mean lexical densities with 95% confidence intervals.

To answer research question 4, we compare the speech rate in our corpora. The literature does not suggest that speech rate varies disciplinarily, which is why we only compare the speech rate of science podcasts against that of the MICASE's mixed-disciplinary lectures. Nesi (2001) and Wingrove (2017) used 5-minute and 1-minute samples from their corpora, respectively, to determine speech rate and did not deduct any pauses, i.e. interruptions in speech. Unlike them, we use complete podcasts and complete lectures to determine speech rate and we also follow the methodology from the seminal paper by Tauroza and Allison (1990) in deducting all silent periods longer than 3 seconds – these are considered interruptions and are excluded from the data. In the MICASE lectures, all pauses longer than 3 seconds are indicated in terms of their duration in the transcripts – these were added up and deducted from the total duration of the lectures (an average deduction per lecture was 146 seconds). For the deduction of the interruptions in science podcasts, the programme Audacity (Audacity Team, 2022) was used. This resulted in slight reductions of the total podcast durations (on average, 10.5 seconds per podcast), as there were not many interruptions longer than 3 seconds in them (typically, podcast audios are processed before publishing and interruptions are deleted). To determine speech rate, we used 21 complete science podcasts, i.e. 7 podcasts per each three podcast sources indicated above, as well as 21 complete MICASE lectures. We express the results in words-per-minute (w.p.m.), as this allows comparison with all previous studies.

4. Results and analysis

The results and the analysis are presented in three sections: vocabulary profile, lexical density and speech rate.

4.1. Vocabulary profile

In this section, we will determine the vocabulary load and the presence of specialised vocabulary in our three corpora.

Table 1 presents the lexical coverages on Nation's BNC/COCA word lists (2012) in the science-podcast corpus, as well as the entire MICASE-lecture

corpus and the MICASE science-lecture sub-corpus. Along with the frequency-based word lists, we use two supplementary lists: that of proper names and that of marginal words. Marginal words encompass all letters (except A and I), typically used as symbols in formulae, equations and measurements (b, c, x, y, Cl, N, O, m, g, l…); these constitute a non-negligible portion of the words in lectures, especially in science lectures (unlike in science podcasts, where they made up only 0.19% of the word-count). These two supplementary word lists were used by Dang and Webb (2014) in their lexical profiling study of the BASE corpus, another well-known lecture corpus, for the same reasons.

BNC/COCA word lists	Science podcasts (cumulative coverage %)	MICASE lectures (cumulative coverage %)	MICASE science lectures (cumulative coverage %)
Proper names (PN)	1.84	2.14	1.21
Marginal words (MW)	2.03	5.22	4.21
1,000 + PW + MW	81.75	85.27	83.81
2,000 + PW + MW	88.44	**91.17**	89.93
3,000 + PW + MW	**93.05**	94.77	**93.81**
4,000 + PW + MW	94.47	**95.93**	**95.33**
5,000 + PW + MW	**95.30**	96.56	96.15
6,000 + PW + MW	95.81	97.00	96.67
7,000 + PW + MW	96.20	97.34	97.08
8,000 + PW + MW	96.63	97.55	97.33
9,000 + PW + MW	96.84	97.69	97.50

Table 1. Vocabulary load (%).

As has been previously said, van Zeeland and Schmitt's (2013) lower listening-comprehension vocabulary threshold (good comprehension for many listeners) is set at a 90%-lexical coverage. In the MICASE mixed-disciplinary corpus, this coverage is achieved with the 2,000 most frequent words of English, proper names and marginal words. However, some more words are needed to obtain this coverage in the corpora of science podcasts and science lectures. More precisely, the 3,000 most frequent words of English, together with proper names and marginal words, cover 93.81% of the words in science lectures, and 93.05% in science podcasts, which suggests a somewhat greater lexical load of science podcasts.

Van Zeeland and Schmitt's (2013) threshold for good comprehension for most listeners is set at 95%-lexical coverage and is met with the 3,000 most frequent English words for speech in general (Nation, 2013). As can be seen from Table 1, this level of comprehension is achieved with the 4,000 most frequent words of English in both the mixed-disciplinary lecture corpus and the science lecture corpus, which corresponds to Dang and Webb's finding

for the BASE corpus (2014). However, the most frequent 5,000 words are needed to achieve this level of comprehension in science podcasts – the same as was found by Coxhead and Walls (2012) for TED talks, another internet-spoken genre.

What our findings suggest, therefore, is that science podcasts are somewhat more lexically demanding than lectures. Another conclusion is that they display more similarity in this respect with science lectures than with mixed-disciplinary lectures.

Table 2 presents our findings regarding the presence of frequent academic and scientific vocabulary in our three corpora.

Specialised word lists	Science podcasts (cumulative coverage %)	MICASE lectures (cumulative coverage %)	MICASE science lectures (cumulative coverage %)
Academic Spoken WL	87.21	86.95	87.28
Hard Science WL	86.22	86.38	87.66
Science List	1.36	1.00	1.89

Table 2. Academic and science vocabulary (%).

The coverages achieved by the ASWL (Dang et al., 2017) are very similar in the three corpora, with an especially remarkable similarity for science podcasts and science lectures (87.21% vs. 87.28%) – again, science podcasts are more similar to this subgroup of lectures than to lectures in general. The similarities identified suggest that science podcasts have a good representation of frequent academic vocabulary found in spoken genres.

Relatively similar coverages in the three corpora are also achieved by the HSWL (Dang, 2018). As expected, this list covered the most in science lectures (as the list itself was derived from a corpus of these). The HSWL fares rather similarly in science podcasts and mixed-disciplinary lectures (86.22% vs. 86.38%).

The most technical of the lists here used, the Science List (Coxhead & Hirsh, 2007) covered the most in science lectures (1.89%), as the most technical and specialised of the three genres here analysed. Its coverage was 28% smaller in science podcasts and 47% smaller in mixed-disciplinary lectures.

The HSWL and the SL indicate that science podcasts are a somewhat poorer source of frequent science vocabulary than science lectures, which could have been expected given that this is a popular science genre and not a science genre *per se*. Still, the results show that science podcasts do possess a good amount of science vocabulary.

4.2. Lexical density

Table 3 presents the results for lexical density.

	Science podcasts	MICASE lectures	MICASE science lectures
Lexical density for the entire corpus (merged)	48.95	48.18	48.47
Mean lexical density per podcast/lecture with 95% CI	49.38 95% CI: ±0.39	49.16 95% CI: ±1.67	50.21 95% CI: ±3.05

Table 3. Lexical density (%).

The three corpora display a greater lexical density than that which is expected for general spoken language – Ure (1971) noted that spoken genres typically have a lexical density lower than 40%. The genres here analysed are institutionalised genres on highly specialised topics, which explains their higher information load. The lexical density results for lectures ranging between 47% and 49%, arrived at by Nesi (2001) and Wingrove (2017) largely correspond to our results here.

The differences in the lexical-density scores in the three corpora analysed are all within less than one-percentage point. We performed a two-tailed independent t-test to see whether there were significant differences between the lexical density of science podcasts and that of MICASE lectures – the t-value was 0.33569 and the p-value was .737491, making the result *not significant* at $p < .05$. Also, we compared the lexical density of science podcasts and science lectures – again, the result is *not significant* at $p < .05$ (the t-value is –0.93091. and the p-value is .353355). The results, therefore, suggest that there is no significant difference between the lexical density of science podcasts and that of lectures.

The least variation in the results is noted in science podcasts, as we can say with 95% confidence that their mean lexical density falls within the range of 49.38 ±0.39%. A small variation in these results suggests that this genre has a rather stable and predictable lexical density, which also recommends it for pedagogical purposes.

4.3. Speech rate

The speech rate of science podcasts and MICASE lectures is presented in Table 4. 95%-confidence intervals are also included.

Speech rate	Science podcasts	MICASE lectures
Words per minute	167.82 w.p.m. 95% CI: ±6.23 w.p.m.	152.98 w.p.m. 95% CI: ±8.79 w.p.m.

Table 4. Speech rate.

A two-tailed t-test for independent means was carried out producing the following: t (40) = –2.69, p = .010145. The result is *significant* at p < .05. The graphic representation of the results with 95%-confidence intervals is given in Figure 2.

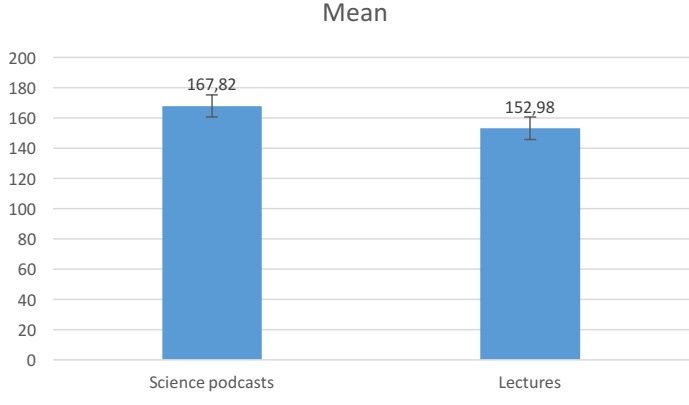

Figure 2. Mean speech rate with 95%-confidence intervals.

The MICASE lectures were delivered at an average rate of about 153 w.p.m., which is a fairly average speech rate, according to both Griffith's (1990) and Tauroza and Allison's (1990) classifications. However, science podcasts are delivered as moderately fast speech – their rate is at the lower end of the moderately fast speech range (165–180 w.m.p., according to Tauroza and Allison (1990)). This speech rate is very similar to that of TED talks (169 w.p.m. (Wingrove, 2017)) and radio broadcasts (160 w.p.m. (Tauroza & Allison (1990)), two genres that share some similarities with podcasts.

With 95% confidence, we may say that the mean speech rate of lectures falls within the range of 144.19–161.77 w.p.m., while that of science podcasts is within the span of 161.59–174.05 w.m.p. We can see that some lectures and podcasts were delivered at a similar speech rate, while amongst some others there may be a difference of as many as 30 w.p.m. Again, there is less variation in the speech rate of science podcasts than lectures, suggesting that this genre has a more predictable speech rate.

Although science podcasts have been found to be faster than lectures, the difference, on average, is about 15 w.p.m., which is certainly not insurmountable. In any case, this potential obstacle may be easily overcome by using various free and simple online tools to slow down the audios. We have downloaded various podcast apps for phones and determined that most of them have in-built options for slowing down and speeding up the podcasts. In addition, it is very easy to replay parts of the audios or whole audios, should that be required. Finally, the mentioned three science-podcast sources have transcripts for most of their podcasts, which can be used for support with lower-proficiency students.

5. Discussion with pedagogical implications

Our findings suggest that science podcasts are somewhat more lexically demanding than lectures. They also display more similarity in this respect with science lectures than with mixed-disciplinary lectures. As we have seen, 1,000 more words are necessary to achieve good comprehension for the genre of science podcasts in relation to mixed-disciplinary lectures and science lectures.

The literature suggests that native speakers expand their vocabularies by about 1,000 words annually until their early twenties (Nation, 2013) and it is considered that L2 language learners have the capacity to match this growth rate provided that they invest a sustained effort in terms of both exposure to input and deliberate vocabulary learning (Nation & Anthony, 2016). Milton (2010) and Capel (2012) assess that a vocabulary size of 5,000 word families or more is that of advanced students (C-levels according to CEFR (Council of Europe, 2001)), which suggests that science podcasts can be followed by advanced learners vocabulary-wise. The same authors suggest that a vocabulary size of 4,000 words is typically that of upper-intermediate learners (B2-level according to CEFR (Council of Europe, 2001)), which suggests that upper-intermediate learners would need vocabulary support in order to follow science podcasts successfully, as was the case with Hawke's (2010) lower-proficiency students. Based on this, we may say that science podcasts are best suited for upper-intermediate and advanced learners.

As for academic vocabulary, we found that the ASWL (Dang et al., 2017) had a very similar coverage in both science podcasts and lectures, which recommends science podcasts as a genre which can provide students with

adequate input in terms of frequent academic vocabulary. Therefore, based on the presence of academic vocabulary, we can say that science podcasts are suitable for EAP, the same as was argued by Ye (2021).

Similar coverages in the three corpora were also achieved by the HSWL (Dang, 2018), while the results for the SL (Coxhead & Hirsh, 2007) – science podcasts had 28% less of this vocabulary compared to science lectures. Thus, science podcasts are a somewhat poorer source of scientific vocabulary than science lectures – still, the amount of this type of vocabulary in them is non-negligible and can recommend this genre for occasional use in EST.

The results for lexical density were extremely similar in all three corpora, which recommends science podcasts for both EAP and EST listening, given that they do not have any greater information load than lectures.

Finally, we noted that science podcasts are faster than lectures. This may pose an obstacle for some learners, but not an insurmountable one, given that the difference in speech rate is not immense. In addition, using in–built options in podcast apps makes it possible to slow down the audios as required (typically, it is sufficient to slow them down by 10%).

It should here be noted that our pedagogical implications do not consider the knowledge of other languages learners may know, particularly their mother tongue. This could be relevant given that EAP and EST, in particular, are rich in words of Latin and Greek origin (Green, 2015). Knowledge of Greek or any Romance language may facilitate listening to science podcasts – learners might be able to interpret the meaning of some cognates. Also, these particular words are typically internationalisms, so those well familiar with them in their mother tongue may require fewer words to follow the genres here analysed.

Regarding limitations, this study was conducted using quantitative methods, based on three criteria, as described earlier. The assessment could have been performed by applying additional qualitative criteria, including studying the feedback of the actual EAP and EST learners listening to science podcasts.

6. Conclusion

In this study, we sought to answer the question of how suitable science podcasts are for EAP and EST listening activities, based on three criteria: their vocabulary profile, lexical density and speech rate.

On the one hand, science podcasts are somewhat more demanding in terms of vocabulary than mixed-disciplinary lectures and science lectures – they can be easily followed by advanced students, while students with lower proficiency would typically need support. On the other hand, science podcasts possess a substantial amount of academic and scientific vocabulary.

The lexical density of science podcasts is very similar to that of mixed-disciplinary lectures and science lectures. Their speech rate, however, is faster, but this potential problem can be overcome quite simply by using the built-in options for slowing down audios in podcast apps or by using other software to adjust the speech rate, should this be necessary.

Based on our analysis, we recommend science podcasts for both EAP and EST listening.

Article history:
Received 16 July 2023
Received in revised form 20 November 2023
Accepted 20 December 2023

References

Abdurakhmonova, D., & Abduvohidova, H. (2023). Challenges in Listening to second language learners. *Central Asian Journal of Literature, Philosophy and Culture*, *4*(3), 74-77.

Al-Ahdal, A.A.M.H. (2020). Overcoming pronunciation hurdles in EFL settings: An evaluation of podcasts as a learning tool at Qassim University Saudi Arabia. *Asian EFL Journal Research Articles*, *27*(1), 86-101.

Anthony, L. (2014). *AntWordProfiler (Version 1.4.1)*. Waseda University. http://www.laurenceanthony.net/software

Audacity Team (2022). *Audacity (Version 3.2.2.0)*. https://www.audacityteam.org/

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253-279. https://doi.org/10.1093/ijl/6.4.253

Benavent, G. T., & Peñamaría, S. S-R. (2011). Use of authentic materials in the ESP classroom. *Encuentro*, *20*, 89-94.

Breeze, R. (2008). Researching simplicity and sophistication in student writing. *International Journal of English Studies*, *8*(1), 51-66. https://doi.org/10.6018/ijes.8.1.49091

Browne, C., Culligan, B., & Phillips, J. (2014). The new academic world list. http://www.newgeneralservicelist.org/nawl–new–academic–word–list/

Brunfaut, T., & Revesz, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, *49*(1), 141-168. https://doi.org/10.1002/tesq.168

Capel, A. (2012). Completing the English vocabulary profile: C1 and C2 vocabulary. *English Profile Journal, 3*(1), 1-14. https://doi.org/10.1017/S2041536212000013

Chaikovska, O., Zbaravska, L., & Bilyk, T. (2019). Podcasts in teaching EFL for students majoring in engineering. In L. Malinovska & V. Osadcuks (Eds.), *Proceedings of 18th International Scientific Conference "Engineering for Rural Development"* (pp. 1915-1920). Latvia University of Life Sciences and Technologies. https://doi.org/10.22616/ERDEV2019.18.N344

Council of Europe (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. CUP.

Coxhead, A. (2000). A new academic word list. *TESOL*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Coxhead, A. (2018). *Vocabulary and English for Specific Purposes research: Quantitative and qualitative perspectives*. Routledge. https://doi.org/10.4324/9781315146478

Coxhead, A., & Hirsh, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *12*(2), 65-78. https://doi.org/10.3917/rfla.122.0065

Coxhead, A., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *TESOLANZ Journal*, *20*(1), 55-67.

Csomay, E., & Petrović, M. (2012). "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System*, *40*, 305-315. https://doi.org/10.1016/j.system.2012.05.004

Dang, T. N. Y. (2018). A hard science spoken word list. *ITL – International Journal of Applied Linguistics*, *169*(1), 44-71. https://doi.org/10.1075/itl.00006.dan

Dang, T. N. Y. (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching disciplines-related TV programs: Insights from medical corpora. *TESOL Quarterly*, *54*(2), 436-459. https://doi.org/10.1002/tesq.552

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66-76. https://doi.org/10.1016/j.esp.2013.08.001

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959-997. https://doi.org/10.1111/lang.12253

Dlott, A. M. (2007). A (pod)cast of thousands. *Educational Leadership*, *64*(7), 80-82.

Fadda, H. A., & Qasim, N. A. (2013). From call to mall: The effectiveness of podcast on EFL higher education students' listening comprehension. *English Language Teaching*, *6*(9), 30-41. https://doi.org/10.5539/elt.v6n9p30

Fadhil, A., Gunawan, W., & Wirza, Y. (2023). Lexical density in EFL Indonesian textbooks: A comparative analysis. *JALL (Journal of Applied Linguistics and Literacy)*, *7*(1), 121-136. http://dx.doi.org/10.25157/jall.v7i1.9727

Fox, A. (2008). Using podcasts in the EFL classroom. *TESL–EJ*, *11*(4), 1-13.

Fujita, K. (2017). On the parallel evolution of syntax and lexicon: A Merge-only view. *Journal of Neurolinguistics*, *43*, 178-192. https://doi.org/10.1016/j.jneuroling.2016.05.001

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics,35*(3), 305-327.

Green, T. M. (2015). *The Greek & Latin roots of English*. Rowman & Littlefield Publishers. https://doi.org/10.1093/applin/amt015

Griffiths, R. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning*, *40*, 311-336. https://doi.org/10.1111/j.1467-1770.1990.tb00666.x

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly*, *26*(2), 385-390. https://doi.org/10.2307/3587015

Halliday, M. A. K. (2002). *On Grammar: Volume 1*. A&C Black.

Halliday, M. A. K., & Hasan, R. (2000). System and text: Making links. *Text & Talk*, *20*(2), 201-210.

Hawke, P. (2010). Using internet-sourced podcasts in independent listening courses: Legal and pedagogical implications. *Jalt CALL Journal*, *6*(3), 219-234. https://doi.org/10.29140/jaltcall.v6n3.102

Hayati, A. M. (2010). Notes on teaching English pronunciation to EFL learners: A case of Iranian high school students. *English Language Teaching*, *3*(4), 121-126. https://doi.org/10.5539/elt.v3n4p121

Hiltunen, T., & Tyrkkö, J. (2018). Academic vocabulary in Wikipedia articles: Frequency and dispersion in uneven datasets. In C. Suhr, T. Nevalainen & I. Taavitsainen (Eds.), *From data to evidence in English language research* (pp. 282-306). Brill. https://doi.org/10.1163/9789004390652_013

Jenkins, J. (2012). English as a lingua franca from the classroom to the classroom. *ELT Journal*, *66*(4), 486-494. https://doi.org/10.1093/elt/ccs040

Jensen, E. D., & Vinther, T. (2003). Exact repetition as input enhancement in second language acquisition. *Language Learning*, *53*(3), 373-428. https://doi.org/10.1111/1467-9922.00230

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University*, *53*, 61-79.

Kavaliauskienė, G. (2008). Podcasting: A tool for improving listening skills. *Teaching English with Technology*, *8*(4), 1-17.

Knight, R. (2010). Sounds for study: Speech and language therapy students: Use and perception of exercise podcasts for phonetics. *International Society for Exploring Teaching and Learning*, *22*,

269-276.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Laurén & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Multilingual Matters.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322. https://doi.org/10.1093/applin/16.3.307

Lee, M. J., & Chan, A. (2007). Pervasive, lifestyle□integrated mobile learning for distance learners: an analysis and unexpected results from a podcasting study. *Open Learning: The Journal of Open, Distance and e-Learning*, *22*(3), 201-218. https://doi.org/10.1080/02680510701619810

Li, H. C. (2012). Using podcasts for learning English: Perceptions of Hong Kong secondary 6 ESL students. *ELT World Online*, *4*, 78-90. https://doi.org/10.21070/jees.v5i2.767

Lindqvist, C., Gudmundson, A., & Bardel, C. (2013). A new approach to measuring lexical sophistication in L2 oral production. In C. Bardel, C. Lindqvist & Laufer, B. (Eds.), *L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis* (pp. 109-126). Eurosla.

Liu, C. Y. (2021). Examining the implementation of academic vocabulary, lexical density, and speech rate features on OpenCourseWare and MOOC lectures. *Interactive Learning Environments*, 1-16. https://doi.org/10.1080/10494820.2021.1987274

Liu, C. Y. (2023). Are TED talks potential materials for learning specialized vocabulary? A case of medical vocabulary. *English Teaching & Learning, 47*(1), 97-118. https://doi.org/10.1007/s42321-021-00105-2

López-Carril, S., Añó, V., & González-Serrano, M. H. (2020). Introducing TED talks as a pedagogical resource in sport management education through YouTube and LinkedIn. *Sustainability*, *12*(23), 10161. https://doi.org/10.3390/su122310161

Lord, G. (2008). Podcasting communities and second language pronunciation. *Foreign Language Annals*, *41*(2), 364-379. https://doi.org/10.1111/j.1944-9720.2008.tb03297.x

Lukmana, I., & Gunawan, W. (2021). Lexical density in UU Cipta Kerja: A case study of controversial articles. In *Proceedings of the Tenth International Conference on Languages and Arts (ICLA 2021)* (pp. 114-119). Atlantis Press.

MacKenzie, L. (2019). Science podcasts: Analysis of global production and output from 2004 to 2018. *Royal Society Open Science*, *6*(1), 180932. https://doi.org/10.1098/rsos.180932

Martínez, I. A. (2011). Capitalizing on the advantages of the Latin American EAP situation: Using authentic and specific materials in EAP writing instruction. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, *21*, 31-47. https://www.revistaiberica.org/index.php/iberica/article/view/327/313

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Bartning, M. Martin & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211-232). Eurosla.

Mulyanti, W., & Soeharto, P. P. (2020). Text complexity in English textbooks for junior high school: a systemic functional perspective. In *Twelfth Conference on Applied Linguistics (CONAPLIN 2019)* (pp. 217-222). Atlantis Press. https://doi.org/10.2991/assehr.k.200406.044

Myers, G. (2003). Discourse studies of scientific popularizations: Questioning the boundaries. *Discourse Studies*, *5*(2), 265-279. https://doi.org/10.1177/14614456030050020

Nagy, William E. (1988). *Teaching vocabulary to improve reading comprehension*. International Reading Association.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, P. (2012). *The BNC/COCA word family lists*. http://www.victoria.ac.nz/lals/about/staff/paul–nation

Nation, P. (2013). *Learning vocabulary in another language* [2nd ed.]. CUP.

Nation, P., & Anthony, L. (2016). Measuring vocabulary size. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 355-368). Routledge.

Nesi, H. (2001). A corpus-based analysis of academic lectures across disciplines. In J. Cotterill, & A. Ife (Eds.), *Language across Boundaries: Selected papers from the annual meeting of the British Association for Applied Linguistics* (pp. 201-218). BAAL & Continuum Press.

O'Bannon, B. W., Lubke, J. K., Beard, J. L., & Britt, V. G. (2011). Using podcasts to replace lecture: Effects on student achievement. *Computers & Education*, *57*(3), 1885-1892. https://doi.org/

10.1016/j.compedu.2011.04.001

Parkinson, J., & Adendorff, R. (2004). The use of popular science articles in teaching scientific literacy. *English for Specific Purposes*, *23*, 379-396. https://doi.org/10.1016/j.esp.2003.11.005

Parkinson, J., Jackson, L., Kirkwood, T., & Padayachee, V. (2007). A scaffolded reading and writing course for foundation level science students. *English for Specific Purposes*, *26*, 443-461. https://doi.org/10.1016/j.esp.2007.01.001

Putman, S. M., & Kingsley, T. (2009). The atoms family: Using podcasts to enhance the development of science vocabulary. *The Reading Teacher*, *63*(2), 100-108. https://doi.org/10.1598/RT.63.2.1

Rader, K. E. (1990). *The effects of three different levels of word rate on the listening comprehension of third-quarter university Spanish students*. The Ohio State University.

Rafique, R. (2019). Using podcasts to improve listening skills of tertiary level learners: Insights from literature. *The Dhaka University Journal of Linguistics*, *11*(21-22), 117-132.

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, *35*(1), 31-65. https://doi.org/10.1017/S0272263112000678

Rolls, H., & Rodgers, M. P. (2017). Science-specific technical vocabulary in science fiction-fantasy texts: A case for 'language through literature'. *English for Specific Purposes*, *48*, 44-56. https://doi.org/10.1016/j.esp.2017.07.002

Schmitt, N., Xiangying, J., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Stanley, G. (2006). Podcasting: Audio on the Internet comes of age. *TESL–EJ*, *9*, 1-7.

Stubbs, M. (1986). Lexical density: A computational technique and some findings. In M. Coulthard (Ed.), *Talking about text* (pp. 27-48). University of Birmingham.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, *11*(1), 90-105. https://doi.org/10.1093/applin/11.1.90

Ure, J. (1971). Lexical density and register differentiation. In G. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443-452). CUP.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457-479. https://doi.org/10.1093/applin/ams074

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

Vuković-Stamatović, M. (2022). Suitability of science & technology documentaries for EAP and EST listening. *Journal of English for Academic Purposes*, *58*, 101137. https://doi.org/10.1016/j.jeap.2022.101137

Wingrove, P. (2017). How suitable are TED talks for academic listening? *Journal of English for Academic Purposes*, *30*, 79-95. https://doi.org/10.1016/j.jeap.2017.10.010

Wu, S. R. (2008). Evaluation of the learning of scientific English in podcasting PCs, MP3s, and MP4s scenarios. In M. Singhal, G. D. M. Seruguendo, J. Tsai, W.-C. Lee, K. Romer, Y.-C. Tseng & H.C.W. Hsiao (Eds.), *2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (sutc 2008)* (pp. 537-542). IEEE. https://doi.org/10.1109/SUTC.2008.90

Ye, Y. (2020). EAP for undergraduate science and engineering students in an EFL context: What should we teach? *Ampersand*, *7*, 100065. https://doi.org/10.1016/j.amper.2020.100065

Ye, Y. (2021). From abstracts to "60-second science" podcasts: reformulation of scientific discourse. *Journal of English for Academic Purposes*, *53*, 101025. https://doi.org/10.1016/j.jeap.2021.101025

Zhang, X. (2022). Are TED talks suitable materials for humanities specialised vocabulary learning? *International Journal of Linguistics, Literature and Translation*, *5*(10), 49-57. https://doi.org/10.32996/ijllt.2022.5.10.6

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics*, *18*(1), 49-68.

Milica Vuković Stamatović 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. She has published in journals such as *Journal of English for Academic Purposes, Corpus Linguistics and Linguistic Theory, Discourse & Society, Pragmatics & Society, Journal of Language and Politics, Gender and Language*, among other journals.

Dragana Čarapić is a senior lecturer engaged at the Faculty of Philology, Faculty of Electrical Engineering, Faculty of Mechanical Engineering, Faculty of Metallurgy and Chemical Engineering and Faculty of Medicine. She is conducting research in the fields of Lexical Semantics (Near-synonymy and Slang), Lexicology, Translation Studies (Gender bias phenomenon of machine translation tools), Methodology, and English for Specific Purposes (Medicine, Biology, Metalurgy, Chemical Engineering, and Mechanical Engineering).