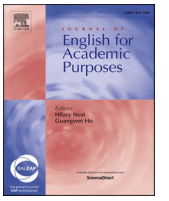
Journal of English for Academic Purposes 58 (2022) 101137

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap



Suitability of science & technology documentaries for EAP and EST listening

Milica Vuković-Stamatović

University of Montenegro, Danila Bojovica bb Niksic, Montenegro

A R T I C L E I N F O

Keywords:
Science and technology documentaries
Lexical level
Academic vocabulary
Technical vocabulary
Lexical density
Speech rate

A B S T R A C T

Selecting English for Academic Purposes (EAP) and English for Science and Technology (EST) listening materials is a challenging task as many different criteria could inform that choice. This study examines the suitability of science and technology documentaries for EAP and EST listening based on a corpus of 100 such documentaries. Departing from the premise that criteria such as thematic suitability, appeal and visual support for the vocabulary presented, as well as the fact that they are delivered in standard English, might recommend science and technology documentaries for this purpose, we further explore their lexical profile, lexical density and speech rate, so as to pass a more informed judgement in this regard. The results obtained are compared against those for academic lectures in general and science lectures in particular. They show that science and technology documentaries may be reasonably followed with a vocabulary size of 3,000 to 4,000 words, similar to what is required for lectures. Additionally, this genre is a good source of academic and technical vocabulary. Its higher lexical density is compensated by its slower delivery than that of lectures. Based on this, we support the inclusion of science and documentary films as EAP and EST materials.

1. Introduction

Many English Language Teaching (ELT) professionals prefer to include authentic teaching and learning materials in their lessons (Reppen, 2010) and such materials are particularly in demand in teaching English for Academic Purposes (EAP) and English for Science and Technology (EST) (cf. Benavent & Peñamaría, 2011). Thus, EAP and EST instructors and material developers often find themselves in a challenging situation of having to select adequate authentic materials (cf. Parkinson, 2000; Parkinson & Adendorff, 2004; Parkinson et al., 2007; Ye, 2020), including those intended for classroom listening exercises or those recommended for extensive listening at home (Takaesu, 2013; Ye, 2021). On the one hand, these materials need to meet certain vocabulary demands (Newton & Nation, 2020). First, their vocabulary must be within the grasp of the target students – a high-level of vocabulary in documentaries used in class has been found to cause frustration among low-level students (Jurkovič, 2013; Swaffar & Vlatten, 1997). In addition, the materials intended for EAP instruction need to cover academic vocabulary to a sufficient degree, whereas those intended for EST classes also need to cover some more specialised, technical vocabulary. On the other hand, when it comes to selecting listening materials specifically, one needs to take into account how lexically dense the delivery in question is, as well as how fast it is paced, given that too much information or speech delivered within a certain timeframe could pose an impediment for some learners (Wingrove, 2017). Moreover, the materials need to be as interesting as possible, as their appeal may contribute to learners' motivation for learning.

E-mail address: vmilica@ucg.ac.me.

https://doi.org/10.1016/j.jeap.2022.101137
Received 29 July 2021; Received in revised form 19 May 2022; Accepted 20 May 2022
Available online 31 May 2022

These are just some of the main criteria one should have in mind when selecting EAP and EST materials, which makes that process not only challenging but also time-consuming (cf. López-Carril, Añó, & González-Serrano, 2020).

Corpus-based investigation into genres which may be included in teaching and learning materials can significantly help ELT professionals and this is what the present study seeks to accomplish – exploring the suitability of one particular genre, science and technology documentaries, which can be considered for use in EAP and EST instruction. As such, this study is situated within a recent and growing line of similar research – some examples of this type of research include Csomay and Petrović's investigation into legal vocabulary contained in discipline-related movies and TV shows to judge their suitability for Legal English classes (2012), Wingrove's study into using TED talks for academic listening (2017), Rolls and Rodger's exploration of the suitability of science fiction-fantasy for ESP classes (2017), Hiltunen and Tyrkkö study of academic vocabulary in Wikipedia articles (2018), Dang's analysis on the presence of specialised vocabulary in medical TV programmes (2020), etc.

The potential of using documentary films in general in ELT is vast (Soong, 2012) and has been explored in several studies such as Rodgers (2018), Peters and Webb (2018), and Peters (2019). Rodgers found that documentaries are a good input for vocabulary learning given that "the same referent is likely to be talked about for a prolonged time" (2018: 205) and that, in them, as much as 65% of the imagery occurs simultaneously with its aural presentation, while some 70% of it occurs in the close proximity of the target words it is associated with and whose learning it may support. These percentages were more than double those Rodgers found for narrative TV genres. This feature of documentaries is especially important in the light of Peters' finding (2019) that words accompanied by on-screen imagery are almost three times more likely to be learned compared to those not accompanied by the supporting imagery. Peters (2019) further argues that this combination of word repetitions and the accompanying imagery that characterises TV documentaries makes them a good source for vocabulary learning. In addition, Peters and Webb find that exposing learners to a single 1-h documentary episode leads to their learning four new words on average and conclude that TV viewing can be "a fruitful method for increasing a learner's vocabulary knowledge" (2018: 24). In addition, documentary films have been found to improve student engagement and learning in ELT (Zakharova, 2017). Not only can a documentary film be motivating and entertaining in terms of its content and presentation, but it could also be "a profoundly engaging tool for raising awareness of vital issues impacting the world today" (Koprowski, 2013, p. 63).

As a subtype of TV documentaries, science and technology documentaries belong to popular science discourse, which is "a key element of academic discourse" (Hyland, 2009, p. 173), given its widespread dissemination and the fact that it exercises a considerable influence on personal and public conceptions, as well as public debates about science. Bearing this in mind, applied linguists have shown an interest in popular science genres as they have sought to improve the teaching of language for academic and specific purposes (Myers, 2003; Parkinson & Adendorff, 2004; Parkinson et al., 2007).

Science and technology documentaries are produced for a range of audiences who do not have a professional need for information about science but want to keep up with the latest scientific developments (Hyland, 2009; 2010). They typically deliver linear, didactic tales, in expository and explanatory modes (Van Dijck, 2006), which focus on "establishing the novelty and relevance of a topic to celebrate scientific results, with their validity taken for granted" (Hyland, 2010, p. 119). They have a strongly narrative storyline, often delivered in the form of a detective story (Curtis, 1994; Dugan, 2014), to make the process of scientific discovery more interesting and entertaining. This is one of the reasons why this genre is very popular – for example, thanks to them, the Discovery Channel has become the most widely distributed cable network in the USA (Hyland, 2009, p. 159).

The above characteristics could make science and technology documentaries good candidates for use in EAP and EST instruction. However, a few additional questions should be addressed before passing a more informed judgement in this regard. These questions include: How lexically demanding are science and technology documentaries? For what English level of students can they be used best? How much specialised, both academic and technical vocabulary, do they contain? How lexically dense are they? Is their pace of delivery suitable for EAP/EST learners? The goal of this study is to answer these questions which are broadly concerned with two issues: that of this genre's vocabulary make-up and that of the information load delivered in them. Based on the answers found to these questions, we will make some pedagogical recommendations regarding their use as EAP and EST listening materials.

2. Background

In this part of the paper we review how the vocabulary profile of texts and speech, as well as specialised vocabulary's representation in them, may be determined using a corpus-based methodology. Additionally, we deal with the issue of determining the information load of speech through two linguistic measures – lexical density and speech rate. We also present some previous studies which have explored the suitability of some genres for ELT purposes.

2.1. Lexical profiling and specialised vocabulary

Lexical profiling, a method and approach first devised by Laufer and Nation (1995), determines the distribution of words in a corpus based on their frequency in reference corpora. Creating such a lexical profile is useful for establishing the vocabulary level of the text analysed, how many words are required for its comprehension, how much specialised vocabulary it contains, how suitable it is for various reading and listening purposes, etc. (Nurmukhamedov & Webb, 2019). The text analysed is loaded into a lexical profiling software, together with a choice of word lists, depending on the goals of the text analysis. Various corpus-based word lists have been produced for lexical profiling research and ELT instruction purposes.

The history of corpus-based word lists dates back as far as 1953, the year in which West manually produced the General Service List (GSL) from his 5-million-word corpus (for details explaining how West's word list (1953) was manually produced from a corpus, see

Gilner (2011)). This list, containing about 2,000 word families, was considered to represent the most frequent words of English, and thus the words which should first be taught in ELT. For many decades it successfully served this purpose, while in recent times newer lists have been produced, based on substantially larger corpora and by employing much more scientific rigour. For instance, Brezina and Gablasova (2015) and Browne et al. (2013) created two such lists. They are both of approximately the same size as the original GSL and are both called the New GSL (NGSL). Expectedly, they outperform the old lists in more recent corpora to a certain degree, according to their creators' reports. Given their size, these lists are mostly intended for EFL/ESL instruction, although measuring their coverages in various corpora can provide indications as to the lexical level of these corpora – generally, the more words of a corpus come from these lists, the lower the lexical demand of the corpus.

Unlike these standalone lists, which comprise only the most frequent vocabulary of English, Nation (2012) produced a set of 25 word lists, representing the most frequent 25,000 word families[1] of English. Additionally, Nation provides four accompanying word lists containing proper names, marginal words, compounds without a hyphen, and abbreviations. This set may be used for instructional purposes but it is also very suitable for use in lexical research due to its comprehensiveness. It is based on two large corpora: the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). The most frequent BNC/COCA 2,000 word families better represent general high-frequency words than the two New-GSL lists referred to above (Dang et al., 2020; Dang & Webb, 2016). The coverage of Nation's word list set may offer useful insights into the lexical make-up of corpora, especially when combined with the knowledge of how much word coverage is needed for reading and listening comprehension. According to Laufer (1989), readers should know at least 95% of the words used in a text for its reasonable comprehension, while Nation (2006, 2013) argues that at least 98% of the words should be familiar to a reader for its ideal comprehension. Van Zeeland and Schmitt (2013) conducted a series of experiments to establish the coverage thresholds for listening comprehension and concluded that 95% coverage is needed for 'adequate' comprehension ("good, but not necessarily complete comprehension" (Van Zeeland & Schmitt, 2013: 475)) achieved by most listeners. Additionally, they also established that 90% coverage led to successful listening comprehension for many experiment participants when listening to "informal narratives", although the scores varied more for this figure than for the one cited above. They also found that 98% coverage, established for ideal reading comprehension, may be too inflated when applied to listening. Subsequent studies (Durbahn et al., 2020; Giordano, 2021) also support the relevance of 90% and 95% coverages for audio and audiovisual input. Knowing how many words are needed to reach comprehension helps determine the level of learners for whom a certain text may be suitable.

Apart from these general-purpose word lists, some researchers have produced more specialised word lists. A series of word lists containing the most frequent academic vocabulary include the Academic Word List (AWL), created by Coxhead in 2000; the New AWL (Browne et al., 2014); the Academic Vocabulary List (AVL), produced by Gardner and Davies (2014); and the Academic Spoken Word List (ASWL), generated by Dang et al. (2017). Dang et al.'s ASWL (2017) is the most relevant for our present research, given that it represents spoken English as delivered in academic settings. This list was derived from a 13-million-word corpus, containing speech transcripts from 24 different academic subjects. It was built 'from scratch', meaning that no group of words was excluded prior to establishing the word list (unlike the AWL and the New AWL, which were derived from the words outside the GSL and new GSL, respectively). The ASWL contains 1,741 word families, which cover 90.13% of the corpus they came from. Dang et al. compared it to the AWL and the AVL, another recent word list representing written academic language, and found that it strongly outperformed them in spoken academic English.

Likewise, a series of word lists which are based on more specialised corpora, typically from one or several closely related disciplines, have emerged over the past few decades. Among them is the EAP Science List (EAP SL) by Coxhead and Hirsh (2007), which was obtained by first excluding the most frequent general and academic words (the GSL and the AWL). This series also includes the word lists for pharmacology (Fraser, 2007), medicine (Lei & Liu, 2016; Wang et al., 2008), engineering (Ward, 2009), chemistry (Valipouri & Nassaji, 2013), environmental science (Liu & Han, 2015), linguistics (Moini & Islamizadeh, 2016), spoken hard science (Dang, 2018), etc. The most relevant of these word lists for our present research is the Hard Science Spoken Word List (HSWL), created by Dang (2018), from a 6.5-million token corpus of transcripts from 12 disciplines. This word list features 1,595 word families which are the most frequent in hard science speech, covering 90.94% of the words in the corpus it was derived from.

All specialised lists can be used to instructional purposes, either as guides for curricula and textbooks or as standalone vocabulary resources, but they could also be used for the purpose of selecting teaching and learning materials. Namely, by checking their coverages in candidate texts we can conclude how valuable these texts are in terms of the vocabulary they contain and how their lexical profile fares against some other relevant genres.

2.2. *Information load*

As suggested by Wingrove (2017), when determining the suitability of some genres for listening, one should also look into their information load, i.e. how lexically dense they are and how much speech is delivered per a unit of time (speech rate).

Lexical density is calculated by dividing the total number of lexical, i.e. content words, by the total number of words in a text (Stubbs, 1986). Lexical words – nouns, verbs, adjectives and adverbs, are differentiated from functional words such as prepositions and articles. Lexical density is used as "an indication of how much information the listener is required to process within a given timeframe" (Wingrove, 2017, p. 81). In his study, Wingrove (2017) finds that the lexical density in the Yale lecture series is 47.11%, similar to that

[1] A word family includes the headword and all its inflected and derived forms (e.g. *create, creates, created, creating, creation, creations, recreate, recreation* …).

he found for his corpus of TED talks – 47.28%. Other results from the literature, *inter alia*, include those from Stubbs (1986), who found a lexical density of 40–65% for non-fiction texts and 40–54% for fiction texts; Ure's finding (1971) that speech generally features a lexical density below 40%; and Nesi's (2001) result for academic lectures (49%).

Speech rate measures how fast speech is delivered per a unit of time, which may be syllables per second or words per minute. Wingrove (2017) reports that the speech rate in his lecture series was 3.52 s.p.s or 144.79 w.p.m., somewhat lower than what he found for TED talks (4.01 s ps or 169 w.p.m). Tauroza and Allison (1990) provide more data for speech rate in various genres: for radio broadcast, the speech rate was 160 w.p.m., for conversation it was 210 w.p.m., for interviews it was 190 w.p.m., while for lectures it was 140 w.p.m., which is comparatively close to Wingrove's finding of 144.79 w.p.m. and Nesi's finding of 150 w.p.m. for the same genre.

Nesi (2001) further studied the relationship between lexical density and speech rate in academic lectures. She found a negative correlation between lexical density and speech rate, meaning that faster speech tended to be less lexically dense and vice versa.

2.3. Suitability of some genres for ELT purposes

A number recent studies have tried to determine the suitability of texts from some particular genres for ELT or, more specifically, for EAP or some EST fields.

Csomay and Petrovíc (2012) explored the specialised, i.e. legal vocabulary, in movies and TV shows on legal topics, so as to judge how suitable they were for use in Legal English teaching and learning. Their results confirmed that these indeed bear a potential for incidental learning of specialised vocabulary given that they contain a solid share of technical, legal terms (6.2%) and that most of them could be encountered more than 10 times in these movies and shows.

As suggested earlier, Wingrove (2017) studied the suitability of TED talks for academic listening. He studied three aspects of his corpus of TED talks in terms of which he compared it against a corpus of academic lectures. First, he studied the presence of academic vocabulary in his corpus of TED talks, using the AVL as a representative word list. Wingrove opted for this list because of its recency, even though it was derived from a corpus of written academic English, bearing in mind that spoken academic word lists were not available at the time of his study. The next aspect studied was the lexical density of his corpus, compared to that of academic lectures. Finally, Wingrove studied the speech rate in his two corpora. It was concluded that TED talks could be used successfully to a wide range of academic listening purposes.

Rolls and Rodgers (2017) examined the coverage of the EAP Science List (Coxhead & Hirsh, 2007) in a corpus of science fiction-fantasy texts, in order to examine whether this genre may be suitable as a source of science-specific vocabulary for EST learners. They found this list to cover 0.50% in a corpus of science fiction-fantasy texts, based on which they concluded that science fiction-fantasy may serve as a "bridge resource" for those studying English as a foreign language who specialise in science.

Hiltunen and Tyrkkö (2018) studied the representation of the AWL (Coxhead, 2000) in Wikipedia articles and concluded that many of them resembled academic prose to a great degree. Using the same word list, the representation of academic vocabulary was also studied by McQuillan (2020), but in a different genre. Namely, McQuillan studied the Harry Potter novels in terms of their representation of academic vocabulary and found that they covered as much as 85% of this type of vocabulary, which recommended them for EAP extensive reading.

Hsu's (2018) study showed that the Voice of America News reading materials are useful for acquiring mid-frequency vocabulary in EFL/ESL learners, as are the transcripts of TED talks (Hsu, 2020). To arrive at these conclusions, Hsu applied Nation's (2012) word list set, referred to above, to measure the vocabulary load of his corpus.

Dang (2020) analysed the presence of specialised vocabulary in medical TV programmes, so as to assess their potential for incidental learning of this type of vocabulary. She first created a Medical Spoken Word List containing 895 word types from a medical lectures and seminars corpus of more than half a million words. The coverage of this list was then tested in a corpus of 37 medical TV programmes. The results pointed to a great potential of this type of programmes for learning specialised vocabulary, particularly if they are watched regularly.

Vuković-Stamatović (2020a) studied the lexical profile of seven physics genres (physics research articles, magazine articles, textbooks, popular books, TED talks, academic lectures, and documentary films) and their suitability for academic listening and reading in English for Physics. For this purpose, the author also used Nation's (2012) word list set, as well as calculated the presence of academic vocabulary (as represented by the AWL) and of the more specialised, scientific vocabulary (as represented by the Science List (Coxhead & Hirsh, 2007)). Similarly, Vuković-Stamatović (2020b) studied the vocabulary profile of science popular books for English for Science classes and found that this genre featured a rather significant amount of academic vocabulary (the coverage of the AWL was 5.32%) and scientific vocabulary (the SL covered 1.51%).

As can be seen, most of the studies described above focused on the vocabulary suitability of certain genres – some of them focused on the vocabulary level needed to reach reading/listening comprehension, while some focused on the presence of specialised vocabulary, academic and/or more technical vocabulary, opting for various word lists to determine its representation in the texts analysed. A number of studies encompassed both these aspects. However, Wingrove's study (2017) proved to be the most comprehensive, as it also studied the information load of the genre in question, a criterion that should not be overlooked when judging the suitability of texts for listening purposes. Bearing all this in mind, it is Wingrove's model that we will apply in this study, with some modifications and additions, which we hope will improve this methodological framework for our present purposes.

3. Aim and research questions

The aim of this paper is to determine how suitable science and technology documentaries are for EAP and EST instruction, respectively. This aim will be fulfilled by answering the following research questions:

1. How many words are needed to reach adequate listening comprehension (as defined by Van Zeeland & Schmitt, 2013) in science and technology documentaries in comparison with academic lectures in general and science lectures in particular?
2. How much specialised vocabulary do science and technology documentaries contain in comparison with academic lectures in general and science lectures in particular?
3. What is the lexical density of science and technology documentaries in comparison with academic lectures in general and science lectures in particular?
4. How similar is the speech rate of science and technology documentaries to that of academic lectures?

The first pair of questions seeks to explore the vocabulary profile of science and technology documentaries, while the second set of questions looks into the information load of this genre. In both these aspects, science and technology documentaries are compared to academic lectures, both the general ones and those on science topics, given that our goal is to explore the suitability of this genre for both EAP and EST purposes.

4. Data and the methodological framework

This subsection of the paper gives the corpora details, as well as the analytical steps and procedures used to answer the set research questions.

4.1. Corpora

The data used in this study are the following:

1. A corpus of 100 recent science and technology documentaries collected for this study. The corpus contains the English captions of these documentaries. In their selection we mostly followed earlier studies (Dang, 2020; Webb & Rodgers, 2009a) and therefore considered two main criteria: the popularity of the documentaries and the availability of English subtitles. The documentaries come from two major and very popular broadcasting companies: half the documentaries (50) are National Geographic (NG) documentaries, comprising a total of 282,599 tokens, while the other half (50) are the British Broadcasting Company (BBC) production, encompassing a total of 330,413 tokens; thus, the entire corpus consists of 613,012 tokens. Some of the documentary titles include: *A day in the Earth life; Alan Turing; How to see a black hole; The secret life of your body-clock; How science changed our world; Human brain: how smart can we get?; Tails, you win: the science of chance; The beginning and end of the universe; How to build a nuclear submarine; Light and dark; Renewable energy; Eugenics: science's greatest scandal; Understanding viruses.* The production years of the documentaries range from 2008 to 2020.
2. The British Academic Spoken English (BASE) corpus[2] of 160 academic lectures, totalling 1,274,913 tokens. We chose this rather than the Michigan Corpus of Academic Spoken English (MICASE) corpus based on the fact that the BASE is the larger of the two and that it consequently had a much larger science lectures subsection. In addition, the speech rate data for 30 lectures from this corpus are already available in the literature (Nesi, 2001).
3. The BASE corpus subsection of 80 academic lectures from the fields of Life and Medical Sciences, and Physical Sciences, which are topic-wise the closest to science and technology documentaries. This BASE corpus' subsection contains 583,434 tokens.

4.2. Data analysis

To answer the first two research questions, we employ the Lexical Frequency Profiling method (Laufer & Nation, 1995), referred to earlier. As explained, the method implies loading word lists into a lexical profiling computer programme, as well as a corpus. The software then provides the coverage of the loaded word lists in the corpus used. The lexical profiling software used here is AntWordProfiler 1.4.1 (Anthony, 2014). The word lists used in this study are the following (all of them were briefly described in section 2.1):

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

[2] The corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi and Paul Thompson. Corpus development was assisted by funding from BALEAP, EURALEX, the British Academy and the Arts and Humanities Research Council.

As regards the first research question, the coverages obtained for Nation's word list set are compared with two listening comprehension thresholds relevant for listening comprehension, arrived at by Van Zeeland and Schmitt (2013): the 90% coverage ensuring adequate comprehension for many listeners and the 95% coverage ensuring adequate comprehension for most listeners. This will enable us to know how many words are needed for these levels of comprehension in science & technology documentaries as opposed to academic lectures in general and science lectures in particular.

The second research question refers to the specialised vocabulary representation in our corpora. To calculate the presence of academic vocabulary, Wingrove (2017) used the AVL, a word list representing a written academic corpus, given that at the time of publishing his results, no spoken academic word list was available. As one is now available (the ASWL, Dang et al., 2017), it only makes sense to use it, as differences between spoken and written academic genres in terms of the level of the specialised vocabulary have been noted in the literature (Vuković-Stamatović, 2020a). As our goals in this study encompass the suitability of science & technology documentaries for EST in particular, we also used Dang's Hard Science Spoken Word List (HSWL) (2018b). This list was built from scratch and it encompasses the most frequent words used in spoken hard science. To further profile some more technical vocabulary which is generally outside the most frequent word bands, we used the EAP Science List (Coxhead & Hirsh, 2007). The limitation of this word list is that comes from written academic language; however, its advantage for our particular purpose is that it was built on top of the most frequent general and academic words (the GSL and the AWL), which is why it is more technical in nature than the lists above.

To answer the third research question, as Wingrove (2017), we used the online software *Analyze my Writing* (http://www.analyzemywriting.com/index.html), which calculates the ratio between lexical (content) words and the total number of words in a corpus. We then compared the lexical density in science and technology documentaries against that from academic lectures in general and science lectures in particular.

Finally, to respond to the fourth research question, we compared the speech rate in science and technology documentaries to that in academic lectures. The BASE lectures recordings were not available online but, fortunately, the speech rate for them has already been reported by Nesi (2001), which is why we use her figures for comparison. Nesi used 5-min samples taken from the middle sections of 30 BASE lectures (after the announcements but before the concluding summaries). We also took the same number of 5-min samples, from the middle part of the documentaries (the samples were randomly taken from somewhere between the 15th and the 35th minute of the documentary). Given that TV documentaries also contain music and speech pauses for viewers to reflect on the content and the imagery, these needed to be subtracted. Following the methodology used in the seminal paper by Tauroza and Allison (1990), we deducted all speechless periods larger than 3 s, i.e. we considered such periods as interruptions and excluded them from the data. Syllables and words were counted using the Syllable Count software (available at https://www.syllablecount.com/), as in Wingrove (2017), and the results are presented in two formats: words per minute (wpm) and syllables per second (sps).

5. Results

5.1. Vocabulary level

As explained earlier, to calculate the vocabulary level of our corpora we used Nation's word list set derived from the combined corpus of BNC and COCA (2012) – the results are reported in Table 1 below. Apart from the word frequency bands, we used two of his supplementary lists: proper names and marginal words. The latter chiefly refers to the letters of the alphabet, which are typically used as symbols in equations, formulae and measurements (x, y, a, C, H, O, m, l, s …), and which thus make up a significant portion of the words in academic lectures (as opposed to our documentaries, where they only make up 0.15% of the words). Both these word lists were used by Dang and Webb (2014) in their lexical profiling study of the BASE corpus, for the same reason.

As explained earlier, 90% coverage ensures adequate comprehension for many listeners (Van Zeeland & Schmitt, 2013). As may be seen in Table 1, as few as 2,000 word families (plus proper nouns and marginal words) are needed to comprehend academic lectures from the BASE corpus, while some more are needed for both science and technology documentaries and the BASE subsection consisting of lectures from life, medical and physical sciences.

Also as indicated earlier, 95% coverage ensures adequate listening comprehension for most listeners (Van Zeeland & Schmitt, 2013). By applying this threshold against the results presented in Table 1, we conclude that the most frequent 4,000 word families are needed for the listening comprehension of science and technology documentary films, and for science lectures, while some fewer words, i.e. 3,000 word families, are needed to understand academic lectures in general. 3,000 word families are generally needed to understand spoken English (Nation, 2006), as well as movies (Webb & Rodgers, 2009b), and our results are, thus, consistent with these when it comes to the vocabulary demands of academic lectures in general; at the same time, however, our results point to some more than average lexical demands which are placed in science spoken genres.

Taking both the above vocabulary coverage thresholds into account, the vocabulary demands of science and technology documentaries, and science lectures are the same. Our results, therefore, speak in favour of using science and technology documentaries as EAP and EST listening materials.

Below is a short extract from our corpus of documentaries, given for illustration purposes – the indices mark the level of words, i.e. "1" marks a word which belongs to the most frequent 1,000 words of English, "2" marks a word from the second most frequent 1,000 words of English, etc. The words outside the first 2,000 words are shown in bold:

Table 1
Vocabulary level (%).

BNC/COCA word lists	Science & tech. documentaries (cum. coverage %)	BASE academic lectures (cum. coverage %)	BASE life, medical & physical sciences lectures (cum. coverage %)
Proper names (PN)	1.6	1.1	0.35
Marginal words (MW)	1.75	4.32	3.71
1,000 + PW + MW	81.17	84.08	83.5
2,000 + PW + MW	88.78	**90.37**	89.49
3,000 + PW + MW	**93.78**	**95.06**	**94.2**
4,000 + PW + MW	**95.78**	96.27	**95.58**
5,000 + PW + MW	96.81	96.92	96.39
6,000 + PW + MW	97.43	97.29	96.79
7,000 + PW + MW	97.89	97.57	97.11
8,000 + PW + MW	98.16	97.77	97.36
9,000 + PW + MW	98.35	97.91	97.54

"We[1]'re[1] about[1] to[1] begin[1] a[1] journey[2] that[1] will[1] take[1] us[1] from[1] the[1] **infinitesimal**[12] to[1] the[1] **infinite**[4], from[1] the[1] dawn[2] of[1] time[1] to[1] the[1] **distant**[3] future[2]. We[1]'ll[1] **explore**[3] **galaxies**[4] and[1] suns[1] and[1] worlds[1], **surf**[5] the[1] **gravity**[3] waves[1] of[1] space[1]-time[1], **encounter**[3] beings[1] that[1] live[1] in[1] fire[1] and[1] ice[1], **explore**[3] the[1] planets[2] of[1] stars[1] that[1] never[1] die[1], discover[1] **atoms**[3] as[1] massive[2] as[1] suns[1] and[1] **universes**[3] smaller[1] than[1] **atoms**[3]. **Cosmos**[7] is[1] also[1] a[1] story[1] about[1] us[1]. It[1]'s[1] the[1] **saga**[6] of[1] how[1] wandering[2] bands[2] of[1] hunters[1] and[1] gatherers[2] found[1] their[1] way[1] to[1] the[1] stars[1], one[1] adventure[2] with[1] many[1] heroes[2].

To[1] make[1] this[1] journey[2], we[1]'ll[1] need[1] imagination[1]. But[1] imagination[1] alone[1] is[1] not[1] enough[1] because[1] the[1] reality[1] of[1] nature[1] is[1] far[1] more[1] wondrous[1] than[1] anything[1] we[1] can[1] imagine[1]. This[1] adventure[2] is[1] made[1] possible[1] by[1] generations[2] of[1] searchers[2] **strictly**[3] **adhering**[4] to[1] a[1] simple[1] set[1] of[1] rules[1]. Test[1] ideas[1] by[1] **experiment**[3] and[1] observation[2]. Build[1] on[1] those[1] ideas[1] that[1] pass[1] the[1] test[1]. **Reject**[3] the[1] ones[1] that[1] fail[2]. Follow[1] the[1] evidence[2] wherever[1] it[1] leads[1] and[1] question[1] everything[1]." (From *Cosmos: A Space-Time Odyssey* (2014)).

As can be seen, most words in the extract come from the most frequent 2,000 words of English. However, 18 out of the extract's 166 words are beyond the first two bands. Four words come from the bands beyond the most frequent 4,000 words of English (*infinitesimal, surf, cosmos,* and *saga*), covering 2.4% of the words in this paragraph – a percentage which, generally, should not put at risk the overall comprehension of this extract.

5.2. Specialised vocabulary

To calculate the level of specialised vocabulary in our corpora, we used the Academic Spoken Word List (Dang et al., 2017), the Hard Science Spoken Word List (Dang, 2018), and the EAP Science List (Coxhead & Hirsh, 2007), as suggested earlier. The results are presented in Table 2 below.

The two spoken word lists suggest that science and technology documentaries are very similar to academic lectures in general and science lectures in particular, in terms of their most frequent vocabulary. Still, they feature somewhat less specialised vocabulary than the lectures, about 1% fewer of such words. The best results of these two lists are scored in life, medical and physical science lectures, which are the most specialised in nature.

The results regarding the coverage of the EAP Science List are very interesting. This list covered almost 4% in written science academic language (Coxhead & Hirsh, 2007) but, expectedly, much less than that in our spoken corpora – Table 2 shows that in academic lectures in general its coverage is 1.01% and that it is significantly larger in science lectures, where it is 1.82%. Our results show that science and technology documentaries are a poorer source of this type of vocabulary than science lectures, but a significantly better one than academic lectures in general. This finding speaks in favour of using more specialised material sources for EST purposes and that sometimes the genre is less important than the topic of a text. Based on their calculation that the EAP Science List covers 0.5% of the words in science fiction-fantasy texts, Rolls and Rodgers (2017) concluded that particular genre could be used as a bridge resource for science students learning English; compared to that, science and documentary films are almost a three times richer source of this type of vocabulary. We may conclude that science and technology documentaries feature a fair amount of specialised vocabulary, which can justify their use for both EAP and EST listening. This relatively high load of specialised words in science and technology documentaries should not pose a significant burden for learners' comprehension given our finding above that knowing some 3,000–4,000 most frequent words of English is sufficient to follow this genre, the significant facilitating role of imagery in documentaries (Rodgers, 2018), and the fact that a considerable proportion of scientific words are, in fact, cognates, which belong to the so-called "International Scientific Vocabulary" (Gove, 1968) (cf. Moss, 1992; Quero, 2015), whose meanings Peters and Webb found the easiest for learners to recognise while listening to a documentary (2018: 21).

Again, for illustration purposes, we quote an extract from our documentary corpus. As the ASWL and the HSWL contain the most frequent words and thus cover almost 9 in every 10 words of our corpora, we will here mark the words contained in the EAP SL (using bold), as the most specific of the three lists used in this subsection, to provide an indication of just how much technical language is

[3] The number of syllables in the samples was counted using https://www.syllablecount.com, following Wingrove (2017).

Table 2
Specialised vocabulary (%).

Specialised word lists	Science & tech. documentaries (cum. coverage %)	BASE academic lectures (cum. coverage %)	BASE life, medical & physical sciences lectures (cum. coverage %)
Academic Spoken WL	86.8	87.7	87.87
Hard Science WL	86.1	86.87	87.71
EAP Science List	1.44	1.01	1.82

present in science and technology documentaries:

> "Plants **absorb** most of the rest of the rainbow, the blue and red photons, and use their energy to power **photosynthesis**. **Photosynthesis** is the process by which plants harness light. They are the bridge between nuclear reactions million miles away and life on Earth. Energy released from nuclear **fusion** reactions in the sun's core heats everything up and shakes **electrons** around, and those **electrons** will **emit** photons, which travel across space for eight minutes, and then hit an **electron** in a chlorophyll **molecule**. But instead of that energy being dissipated away as heat, chlorophyll is clever, and ultimately the energy imparted to that **electron** is used to do all sorts of clever things through an intricate piece of machinery. Split water up, force **electrons** onto **carbon** dioxide, and ultimately, build sugars, which allows the plant to grow." (From *Forces of Nature with Brian Cox* (2016))

The science words used in the extract include words such as *electrons* and *molecules*. However, we must bear in mind that specialised word lists typically cover only relatively frequent vocabulary and so is the case with the EAP SL – in the paragraph above, certainly some more words could be profiled as scientific (e.g. *chlorophyll* and *photons*). This means that even more technical words may be found in the genre of science and technology documentaries, which could further recommend it for EST listening.

5.3. Lexical density

As argued by Wingrove (2017), lexical density is one of the indicators of the information load of texts. Generally, the literature cites that spoken texts are typically less dense than written ones (Halliday, 1989). Our results for the density of the corpora here analysed are given in Table 3 below.

Science and technology documentaries are much lexically denser than lectures, featuring a density of 51.36%, which is well above what is expected of spoken texts – as discussed earlier, Ure (1971) found that spoken texts generally have a lexical density lower than 40%. However, we must bear in mind that documentaries are scripted talks – basically a genre which is written for speaking, which is why it is significantly different from spontaneous speech in many respects. Academic lectures are also lexically denser than everyday spoken genres, being less spontaneous than them – although lectures are not scripted *per se*, they are, in general, very well prepared in advance.

The lexical density of science and technology documentaries may render them more difficult to understand than spontaneous speech; however, bearing in mind that their vocabulary load is suitable for English learners with a vocabulary size of 4,000 words and the fact that the images strongly support the delivery in this genre (Rodgers, 2018), this rather large lexical density should not represent an insurmountable obstacle for learners. Other factors related to learners themselves may also help them offset higher lexical density, such as their having a technical background or interest, or their familiarity with the topics of the documentaries; additionally, an important factor may also be their first language or any additional foreign languages they may speak. This may be of relevance as the language of science is rich in words of Greek and Latin origin (Green, 2015). Because of this, learners speaking Romance languages or, in general, those well versed in vocabulary of Greek and Latin origin, might have a vocabulary advantage when listening to science and technology documentaries – they could be able to recognise or guess the meaning of certain cognates. This advantage would, presumably, be greater when reading documentary transcripts or subtitles rather than just listening to them. Finally, if the documentaries are intended for extensive listening outside the classroom, learners can also take advantage of the play-speed and rewind control functions of their playing devices in order to slow down the recording or repeat some of its segments. Naturally, the genre of science and technology documentaries, as any other genre, may be more suitable for some learners, while less for the others, bearing in mind factors such as the above.

Below is another extract from our corpus, with the lexical words marked in bold, so as to provide an indication of their presence in the genre of science and technology documentaries:

Table 3
Lexical density (%).

	Science & tech. documentaries (%)	BASE academic lectures (%)	BASE life, medical & physical sciences lectures (%)
Lexical density(content words/total words)	51.36	46.18	45.88

> "**Animals travel** across the **earth** in **ever-changing formations**. **Landscapes** are **constantly altering**. **One** of the **reasons mathematics began** was because we **needed** to **find** a **way** of **making sense** of these **natural patterns**. The **most basic concepts** of **maths** – **space** and **quantity** – are **hard-wired** into our **brains**. **Even animals** have a **sense** of **distance** and **number**, **assessing** when their **pack** is **outnumbered**, and whether to **fight** or **fly**, **calculating** whether their **prey** is within **striking distance**. **Understanding maths** is the **difference** between **life** and **death**." (From *The Story of Maths* (2008))

The lexical density of the paragraph quoted above is 54.12%, which makes it relatively representative for the corpus. Among the content words, nouns are the most dominant in the extract – they make up 32.94% or about one third of the paragraph, while the remaining content words (adjectives, adverbs and verbs) make up about a fifth of the paragraph.

5.4. Speech rate

Our final research question refers to the speech rate in science and technology documentaries, especially in relation to that in academic lectures. The related findings are in Table 4, along with Nesi's (2001) findings for BASE lectures.

The speech rate in science and documentary films (138.17 w.p.m., SD 17.08, or 3.39 s.p.s.,[3] SD 0.39) is slower than that of academic lectures. The speech rate mean in our corpus is between 132 w.p.m. and 145 w.p.m., based on 30 samples, which suggests some variation, but not that substantial; Nesi (2001) reported much more variation for academic lectures. This scripted genre is typically carefully paced, giving readers time to absorb what is said – a somewhat slower speech rate certainly makes this genre easier to follow.

Our finding is consistent with Tauroza and Allison's (1990) observation that scripted genres have a slower speech rate than spontaneous speech. Nesi's (2001) result that speech rate negatively correlates with lexical density has also been confirmed here: a rather high lexical density of science and technology documentaries is accompanied by their somewhat slower speech rate.

In response to the fourth research question we may say that the speech rate of science and technology documentaries should pose no obstacle for English language learners.

6. Summary

In this paper, we explored a corpus of 100 science and technology documentaries by comparing their lexical profile, lexical density and speech rate, to those of academic lectures in general and science lectures in particular. The purpose of this comparison was to assess the suitability of this genre for EAP and EST listening. For our comparison corpus we used BASE lectures and its science lectures subsection. The vocabulary load of science and technology documentaries proved very similar to them – 3,000 words were found to be necessary for adequate comprehension of this genre for many learners, while 4,000 words were needed to achieve this level of comprehension for most learners. Next, this genre proved to be relatively rich in specialised academic and more technical vocabulary. Its speech rate was somewhat slower than that found for lectures. Finally, the genre turned out to be rather lexically dense.

7. Pedagogical recommendations

Based on the results presented above, science and technology documentaries can be recommended for both EAP and EST listening, as they display great similarities in their lexical profile to academic lectures in general and science lectures in particular, both in terms of their vocabulary level, but also in terms of the presence of the more specialised vocabulary. If the vocabulary level of the students is not familiar to the instructor prior to commencing using this genre for listening materials, we can recommend that they test their learners' vocabulary sizes. This is possible using the Vocabulary Levels Tests (originally created by Nation (1983)), many of which are available online. As suggested by the findings of this study, science and technology documentaries may not be suitable for many learners who do not have sufficient knowledge of the most frequent 3,000 words of English. Also, as mentioned earlier, speakers of Greek or Romance languages could have a vocabulary advantage as scientific English abounds in terminology originating from these languages.

The instructors and the listening material developers should bear in mind that science and documentary films are a scripted genre and as such they feature a rather high lexical density, which is why playing them as videos is recommended – the images facilitate understanding (Peters, 2019; Rodgers, 2018), as does their somewhat slower speech rate, which was here found suitable for language learners.

Some other recommendations as to the specifics on how to use TV documentaries for English vocabulary learning purposes may be found in the literature. Thus, they may be recommended both for classroom listening exercises (typically in the form of excerpts, as shown by Peters (2019), who found vocabulary gains from exposing learners to 12-min documentary excerpts), or for extensive listening at home (the value of which was shown by Peters and Webb (2018), who exposed students to a whole documentary episode). Moreover, Peters (2019) suggests playing them with captions, to ensure best vocabulary gains. Further research is needed, however, to

Table 4
Speech rate (%).

Speech rate	Science & tech. documentaries (no. 30)	BASE academic lectures (no. 30) (Nesi, 2001)	BASE science lectures (no. 10) Nesi (2001)
Words per minute	138.17	150	146

provide more detailed recommendations as to how to best integrate this genre in EAP and EST materials, respectively.

8. Conclusion

The vocabulary profile and the speech rate of science and technology documentaries recommend them for use as EAP and EST listening materials. However, instructors and material developers need to bear in mind that science and technology documentaries are rather lexically dense.

Some limitations of this study should be addressed at this point as well. Namely, the present study was performed using quantitative methods. The suitability of science and technology documentaries for EAP and EST listening may also be assessed using other methods, including the analysis of lexical bundles or by using them in class and surveying both teachers and learners on their suitability or testing how learners coped with them. As suggested earlier, this genre could be more suitable for some learners and less for others, depending on personal characteristics of learners such as their having a technical background or an interest in science and technology, their being familiar with the documentaries' topics, or their speaking other languages, especially Greek or Romance languages. The role of factors such as these in judging the suitability of science and technology documentaries for particular groups of learners could be further explored in future research.

Despite the said limitations, taking into account the previous findings regarding the use of documentaries for the English language teaching and learning (Peters, 2019; Peters & Webb, 2018; Rodgers, 2018) and considering the findings of our study, we support the inclusion of science and documentary films in both EAP and EST listening materials.

Author statement

As I am the only author of this paper, all roles and credits in the process of conceptualising, writing and revising the paper belong to me.

The author.

Milica Vuković-Stamatović.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jeap.2022.101137.

References

Benavent, G. T., & Peñamaría, S. S.-R. (2011). Use of authentic materials in the ESP classroom. *Encuentro, 20*, 89–94. http://hdl.handle.net/10017/10109.

Brezina, V., & Gablasova, G. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Browne, C., Culligan, B., & Phillips, J. (2013). The new general service list. http://www.newgeneralservicelist.org/. (Accessed 1 September 2019).

Browne, C., Culligan, B., & Phillips, J. (2014). The new academic world list. http://www.newgeneralservicelist.org/nawl-new-academic-word-list/. (Accessed 1 September 2019).

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

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

Csomay, E., & Petrovíc, 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

Curtis, R. (1994). Narrative form and normative force: Baconian story-telling in popular science. *Social Studies of Science, 24*(3), 419–461. https://doi.org/10.1177/030631279402400301

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

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

Dang, T. N. Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL - International Journal of Applied Linguistics, 167*(2), 132–158. https://doi.org/10.1075/itl.167.2.02dan

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*. https://doi.org/10.1177/1362168820911189

Dugan, D. (2014). Science story telling in TV documentaries. *Actes d'Història de la Ciència i de la Tecnica*, 35–49. https://doi.org/10.2436/20.2006.01.183

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System, 88*, 1–13. https://doi.org/10.1016/j.system.2019.102166

Fraser, S. (2007). Providing ESP learners with the vocabulary they need: Corpora and the creation of specialized word lists. *Hiroshima Studies in Language and Language Education, 10*, 127–143.

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

Gilner, L. (2011). A primer on the general service list. *Reading in a Foreign Language, 23*(1), 65–83. https://doi.org/10.1057/9781137276285_4

Giordano, M. J. (2021). Lexical coverage in dialogue listening. *Language Teaching Research*, 1–28. https://doi.org/10.1177/1362168821989869

Green, T. M. (2015). *The Greek & Latin roots of English*. Rowman & Littlefield.

Halliday, M. (1989). *Spoken and written language*. OUP.

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

Hsu, W. (2018). Voice of America news as voluminous reading material for mid-frequency vocabulary learning. *RELC Journal, 50*(3), 408–421. https://doi.org/10.1177/0033688218764460

Hsu, W. (2020). Can TED talk transcripts serve as extensive reading material for mid-frequency vocabulary learning? *TEFLIN Journal, 31*(2), 181–203. https://doi.org/10.15639/teflinjournal.v31i2/181-203
Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.
Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes, 9*, 116–127. https://doi.org/10.1016/j.jeap.2010.02.003
Jurkovič, V. (2013). More mighty ships!" Using documentaries in teaching maritime English – student perspectives. In T. Albayrak (Ed.), *International maritime English conference: Proceedings* (pp. 200–217). Piri Reis University. https://doi.org/10.13140/2.1.2267.6167.
Koprowski, M. (2013). Building the four skills through documentary films. In R. Chartrand, et al. (Eds.), *The 2013 PanSIG conference proceedings* (pp. 63–71). JALT PanSIG.
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
Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *English for Academic Purposes, 22*(1), 42–53. https://doi.org/10.1016/j.jeap.2016.01.008
Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes, 39*, 1–11. https://doi.org/10.1016/j.esp.2015.03.001
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), Article 10161. https://doi.org/10.3390/su122310161
McQuillan, J. (2020). Harry potter and the prisoners of vocabulary instruction: Acquiring academic language at hogwarts. *Reading in a Foreign Language, 32*(2), 122–142. http://hdl.handle.net/10125/67377.
Moini, R., & Islamizadeh, Z. (2016). Do we need discipline-specific academic word lists? Linguistics academic word list (LAWL). *Journal of Teaching Language Skills, 35*(3), 65–90. https://doi.org/10.22099/jtls.2016.3901
Moss, G. (1992). Cognate recognition: Its importance in teaching ESP reading courses to Spanish speakers. *English for Specific Purposes, 11*(2), 141–158. https://doi.org/10.1016/S0889-4906(05)80005-5
Myers, G. (2003). Discourse studies of scientific popularizations: Questioning the boundaries. *Discourse Studies, 5*(2), 265–279. https://doi.org/10.1177/1461445603005002006.
Nation, P. (1983). *Testing and teaching vocabulary. Guidelines* (Vol. 5, pp. 12–25).
Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 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. (Accessed 1 April 2018).
Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.) (Cambridge).
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.
Newton, J. M., & Nation, P. (2020). *Teaching ESL/EFL listening and speaking*. Routledge.
Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching, 52*(2), 188–200. https://doi.org/10.1017/S0261444819000028
Parkinson, J. (2000). Acquiring scientific literacy through content and genre: A theme based language course for science students. *English for Specific Purposes, 19*, 369–387. https://doi.org/10.1016/S0889-4906(99)00012-5
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
Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *Tesol Quarterly, 53*(4), 1008–1032. https://doi.org/10.1002/tesq.531
Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through watching a single episode of L2 television. *Studies in Second Language Acquisition, 40*, 551–577. https://doi.org/10.1017/S0272263117000407
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]. Victoria University of Wellington Archive.
Reppen, R. (2010). *Using corpora in the language classroom*. CUP.
Rodgers, M. (2018). The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics, 69*(1), 191–211. https://doi.org/10.1075/itl.00012.rod
Rolls, H., & Rodgers, M. (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
Soong, D. (2012). Using documentary films in oral interpretation class what is the appropriate length? *International Journal of Applied Linguistics & English Literature, 1*(6), 131–141. https://doi.org/10.7575/ijalel.v.1n.6p.131
Stubbs, M. (1986). Lexical density: A computational technique and some findings. In M. Coulthard (Ed.), *Talking about text* (pp. 27–48). University of Birmingham.
Swaffar, J., & Vlatten, A. (1997). Sequential model for video viewing in the foreign language classroom. *The Modern Language Journal, 81*(2), 175–188. https://doi.org/10.2307/328785
Takaesu, A. (2013). TED talks as an extensive listening resource for EAP students. *Language Education in Asia, 4*(2), 150–162. https://doi.org/10.5746/LEiA/13/V4/I2/A05/Takaesu
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.
Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248–263. https://doi.org/10.1016/j.jeap.2013.07.001
Van Dijck, J. (2006). Picturizing science: The science documentary as multimedia spectacle. *International Journal of Cultural Studies, 9*(1), 5–24. https://doi.org/10.1177/1367877906061162
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. (2020a). 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. (2020b). Vocabulary suitability of popular science books for English for science classes: A corpus-based study. In B. Čubrović (Ed.), *Vol. 1. Belgrade English Language and literature studies BELLS90 proceedings* (pp. 439–447). https://doi.org/10.18485/bells90.2020.1.ch26. Faculty of Philology in Belgrade.
Wang, J., Liang, S., & Guang-chun, G. (2008). Establishment of a medical word list. *English for Specific Purposes, 27*, 442–458. https://doi.org/10.1016/j.esp.2008.05.003
Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*(3), 170–182. https://doi.org/10.1016/j.sbspro.2013.01.187
Webb, S., & Rodgers, M. (2009a). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S., & Rodgers, M. (2009b). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427. https://doi.org/10.1093/applin/amp010
West, M. (1953). *A general service list of English words*. Longman.
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
Ye, Y. (2020). EAP for undergraduate science and engineering students in an EFL context: What should we teach? *Ampersand, 7*, Article 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*, Article 101025. https://doi.org/10.1016/j.jeap.2021.101025
Zakharova, M. (2017). National Geographic learning and content-based instruction of English for aviation. In L. Gómez Chova, et al. (Eds.), *EDULEARN17 proceedings* (pp. 1190–1195). IATED Academy. https://doi.org/10.21125/edulearn.2017.1249.

Milica Vuković Stamatović, *PhD, is an Associate Professor of English Language and Linguistics at the Faculty of Philology, University of Montenegro, where she has worked for 15 years. She currently teaches Discourse Analysis, Pragmatics, as well as English for Specific Purposes (English for Science, English for Physics, English for Mathematics, and English for Biology). Her main research interests lie in these three areas.*

https://scholar.google.com/citations?user=9F5SWo8AAAAJ&hl=sr&oi=ao