CAMBRIDGE
UNIVERSITY PRESS

REPLICATION RESEARCH

Examining lexical profile in general-audience English podcasts: A close replication of Nurmukhamedov and Sharakhimov (2021)

Hong Yu[1] (iD) and Ju Wen[2] (iD)

[1]Southwest Petroleum University, Chengdu, People's Republic of China and [2]Chengdu Jincheng College, Chengdu, People's Republic of China
Corresponding author: Ju Wen; Email: ucet@qq.com

Abstract

We closely replicated Nurmukhamedov and Sharakhimov (2021), which was the first study to examine the lexical profile of general-audience English podcasts. Nurmukhamedov and Sharakhimov (2021) found that podcast listeners should have a knowledge of the most frequent 3,000 word families and 5,000 word families, respectively, plus proper nouns, marginal words, transparent compounds, and acronyms in order to achieve good and high-level listening comprehension. We followed the methods and procedures of the initial study with a much larger corpus. Specifically, a total of 8,862 podcast transcripts sampled from 12 general-audience podcast programs were used to compile the 14-million-word corpus. Results of the study confirmed Nurmukhamedov and Sharakhimov's (2021) findings in the vocabulary size required to understand general-audience English podcasts. However, some minor differences pertaining to individual podcast programs were revealed, indicating that the sampling of data had an effect on the lexical demand. These findings provide solid evidence to support the validity and generalizability of the initial study's findings. Implications for second language teaching and learning are also discussed.

1. Introduction

Replication, as a method of validating research in sciences, has received increasing attention in the study of applied linguistics over the past decades (Porte, 2012; Porte & McManus, 2019). Researchers in the field have reached a general consensus on the value of verifying, consolidating, and generalizing the findings reported in established studies with replication (McManus, 2022). As an integral part of applied linguistics, corpus-based studies have been advocated for replication (Egbert & Baker, 2016; Omidian et al., 2023; Stubbs, 2001).

One line of research in corpus-based studies that calls for replication is lexical coverage research. For example, Schmitt et al. (2017) proposed the replication of studies on lexical coverage (van Zeeland & Schmitt, 2013) and vocabulary size (Nation, 2006), aiming to develop "a more reliable, nuanced, and ecologically valid understanding of the amount of vocabulary learners need to acquire in order to become proficient language users in their chosen domain" (Schmitt et al., 2017, p. 214). The reasons that they made such a proposal are twofold: (1) for pedagogical purposes, good estimates of the vocabulary size are crucial for language teaching and learning in that they form learning targets for language learners, and (2) for research purposes, "there are a limited number of studies informing these essential size targets" (Schmitt et al., 2017, p. 214). Thus, replication studies are critically needed in order to clarify the key coverage and size figures.

To support that proposal, Schmitt et al. (2017) suggested two directions for conducting replication in lexical coverage research. The first direction is to increase the corpus size. Findings derived from

"small data sets" need to be "checked with larger, more comprehensive corpora" (Schmitt et al., 2017, p. 217). The second direction is to update the research methodology, such as the word lists and the counting unit. For instance, as Nation's (2018) BNC (British National Corpus)/COCA (Corpus of Contemporary American English) word family lists have been widely employed in the field of vocabulary studies as "a better indication of word frequency" (Schmitt et al., 2017, p. 218), those studies using Nation's (2006) BNC word family lists to lexically profile a particular domain (e.g., Nurmukhamedov, 2017; Tegge, 2017; Webb & Rodgers, 2009a, 2009b) can barely serve as a guideline and are thus ripe for replication. In response to the call of Schmitt et al. (2017), the present study intends to carry out a replication in lexical coverage research by following one of the directions they recommended.

1.1 General background

Lexical coverage refers to the degree to which the running words in the given text(s) are known by readers or listeners (Nation, 2006; Webb, 2021). Previous studies (Hu & Nation, 2000; Laufer, 1989; Schmitt et al., 2011) have established that a lexical coverage of 90%–98% is needed for second language (L2) learners to achieve adequate comprehension, depending on the modality (e.g., spoken/written), the type of texts, and how "adequate" is defined. For example, Laufer and Ravenhorst-Kalovski (2010) suggested 95% and 98% coverage for minimal and optimal reading comprehension, while van Zeeland and Schmitt (2013) proposed 90%, 95%, and 98% coverage for acceptable, good, and high-level listening comprehension, respectively. These coverage percentages are important because they "indicate the vocabulary size necessary for comprehension of a text" (Rodgers & Webb, 2016, p. 165). According to Nation (2006), if 98% coverage is established, then knowledge of 8,000–9,000 word families is required for unassisted comprehension of a written text for general purposes (e.g., novels, news reports, graded reading texts) and 6,000–7,000 word families for a spoken text (e.g., children's movies, unscripted spoken English).

Research on lexical coverage and vocabulary size has important implications for English as a foreign language (EFL) teaching and learning. On the one hand, it reveals the frequency distribution of words in sets of 1,000 word families, indicating which words are high-frequency (i.e., the most frequent 3,000 word families), mid-frequency (i.e., 6,000 word families from the fourth to the ninth 1,000-level), and low-frequency words (i.e., the tenth 1,000 and above) (Nation, 2022). The high-/mid-/low-frequency distinction informs both L2 teachers and learners of the value of "learning vocabulary in relation to frequency levels" and "learning the most frequent words to facilitate comprehension" (Webb, 2021, p. 283). On the other hand, it provides "an indication of the difficulty level of a text" (Nurmukhamedov & Webb, 2019, p. 188) by revealing the vocabulary size required for unassisted comprehension. Accordingly, L2 teachers and learners can select level-appropriate materials and set concrete vocabulary learning targets. Therefore, it is pedagogically crucial to "get these (size) figures right for a variety of text modalities, genres and conditions of reading and listening" (Schmitt et al., 2017, p. 212).

In recent years, podcasts have gained enormous popularity as they may offer a number of benefits to language teaching and learning (see Abraham & Williams, 2009; Facer et al., 2009). The first potential benefit relates to the universal availability of podcast programs. That is, L2 teachers and learners are free to select from the large quantities of podcast programs readily available on the internet, which cover a wide range of topics. Second, podcasts make the ubiquitous learning of language possible and feasible. L2 learners can enjoy "a wealth of authentic, free and easily accessible aural input" (Liu, 2023, p. 20) anytime, anywhere, and at any pace. Last, the manually-checked transcripts offered by some podcast hosting platforms can serve as an extra aid for those who would like to refer to written texts (Nurmukhamedov & Sadler, 2011). Taken together, podcasts can be deemed as a powerful and effective tool for EFL teaching and learning.

Despite the potential benefits that podcasts may provide to L2 teaching and learning, only a limited number of studies have been carried out to date to assess podcasts as a potential teaching material from a lexical perspective (e.g., Liu, 2023; Motamedynia & Shahri, 2022; Nurmukhamedov & Sharakhimov, 2021). While these studies have provided valuable insights into the lexical demands of podcasts, it

remains unknown to what extent the vocabulary size figures revealed are generalizable (see Section 2, 'Motivation for replication', for a detailed explanation). That said, more empirical studies are needed to evaluate the lexical demand of podcasts if they are to be extensively used as a resource for L2 teaching and learning. Therefore, a replication study in this regard is necessary to test if the lexical demand figures still hold in different samples of podcasts. It is hoped that findings obtained from this replication study can inform future L2 pedagogy and teaching material development by providing a reliable indicator of how lexically demanding podcasts are as a material for EFL teaching and learning.

1.2 The initial study

Nurmukhamedov and Sharakhimov's (2021) study was the first that examined the vocabulary size necessary for adequate comprehension of general-audience English podcasts, and has been cited as the foundational study in research on lexical coverage of podcasts. In this study, transcripts of 170 podcast episodes sampled from nine general-audience English podcast programs were used to compile the 1,137,163-word corpus. Results of the study found that podcast listeners need to know the most frequent 3,000 word families plus proper nouns (PN), marginal words (MW), transparent compounds (TC), and acronyms (AC) to reach 96.75% coverage, and the most frequent 5,000 word families (plus PN, MW, TC, and AC) to reach 98.26% coverage. It was also found that there was variation in coverage among podcast programs. Specifically, the vocabulary necessary to gain 95% coverage was consistent among most of the podcasts (i.e., 3,000 word families). In contrast, the vocabulary necessary to gain 98% coverage varied considerably, ranging from 4,000–6,000 word families.

Some methodological considerations in Nurmukhamedov and Sharakhimov (2021) are worth noting. First, the researchers used 95% and 98% instead of 90% coverage to approximate adequate comprehension. They recommended 98% coverage because podcasts generally did not contain "visual clues or subtitles/captions" and therefore listeners had to "depend on their listening skills and vocabulary knowledge" for adequate comprehension (Nurmukhamedov & Sharakhimov, 2021, p. 11). Thus, a higher coverage figure like 98% is more appropriate for successful listening comprehension without visual support. Second, they used "the largest available lists of word families" (Nation, 2016, p. XII) – that is, the BNC/COCA word family lists. Updated from the original BNC word family lists by adding the COCA frequency information, the newly-combined BNC/COCA word family lists featured better generalizability and applicability in both British and American contexts (Schmitt et al., 2017). Moreover, the inclusion of five additional lists (i.e., PN, WM, TC, AC, and *Not in the lists*) differed from previous research, where only two or three additional lists were used (e.g., Nurmukhamedov, 2017; Webb & Rodgers, 2009a, 2009b). Lastly, they used *AntWordProfiler* (Anthony, 2023), which is probably "the best program for using the lists for vocabulary analysis" (Nation, 2020, p. 2). Nation (2020) encouraged the use of *AntWordProfiler* because the *Range* program has not been updated for many years, whereas *AntWordProfiler* is a better-supported and fully functional solution.

2. Motivation for replication

According to Porte and McManus (2019), a close replication revisits a specific study by modifying only one major variable of interest while keeping the remaining variables unchanged. As Schmitt et al. (2017) suggested in their replication proposal, variables that could be usefully manipulated included the corpus size, the word lists, and the counting unit (e.g., lemma). Among the domains investigated, we specifically focused on podcasts and chose to closely replicate Nurmukhamedov and Sharakhimov (2021), with the sampling data/corpus size modified. The motivations for this choice are as follows.

Podcasts, as one of the most compelling listening materials, have maintained a strong growth momentum, despite the COVID-19 pandemic hit (Quah, 2021). The wave of lockdowns and quarantines impacted people's listening behavior and witnessed the strong performance of podcasting as a global industry (Rowe, 2020). Statistics indicated that people listened to podcasts for approximately 4 h per week in 2020 (Goetzen, 2020), increasing to over 6 h per week in 2023 (Whitner, 2023).

Concurrently, the number of podcast episodes soared from over 30 million in 2022 (see Liu, 2023) to 70 million in 2023 (Whitner, 2023). The meteoric rise and huge popularity of podcasts create considerable potentials and opportunities for L2 learning and teaching.

In addition, podcasts are considered as an effective language-learning tool in EFL settings (Facer & Abdous, 2011; Facer et al., 2009). Podcasts have been used to hone students' pronunciation, reinforce speaking strategies, promote listening comprehension, develop intercultural competence, and enhance students' vocabulary learning (Ducate & Lomicka, 2009; Fouz-González, 2019; Liu, 2023; McBride, 2009; Saeedakhtar et al., 2021). Replication-based studies, therefore, are needed to verify the suitability (in this case, lexical demand) of podcasts as pedagogically useful materials.

Furthermore, several possible limitations of the initial study warranted its replication. The first issue is pertinent to the size of the corpus. It should be noted that no consensus has been reached to date on the "ideal" or "adequate" size of corpus for linguistic studies (McEnery & Brookes, 2022). The corpus size in the initial study (i.e., 1,137,163 words) was deemed as relatively small for two reasons. On the one hand, as noted earlier in this section, given the abundance of podcast programs, the vast quantities of podcast episodes, and their varied lengths and topics, a corpus containing only 170 episodes of podcast transcripts may not be large enough to "capture enough of the language for accurate representation" (Reppen, 2022, p. 14). On the other hand, as most podcast providers have made the transcripts of their programs publicly available on the internet, it has become increasingly feasible for users (e.g., teachers, learners, and researchers) to obtain these texts. More importantly, these programs have been carefully transcribed and the transcripts manually checked by professionals in the industry, which has ensured the quality of texts. In other words, the wide availability of high-quality transcripts has made it ideal and practical to create a large corpus that can better represent podcast programs, in general. For example, a 9.6-million-word corpus was created in Liu (2023) to assess the lexical demand (and suitability) of academic podcasts for English for academic purposes. Against such a backdrop, it is reasonable to consider the corpus size of the initial study (i.e., 1.1-millon words) as relatively small.

The second issue in the study is related to the number of episodes sampled from each podcast program. Although there are no hard rules for achieving representativeness in corpus-based studies (Ädel, 2020), it remains unclear to what extent a limited number of episodes (e.g., 20 episodes for each program) can represent a podcast program as a whole. For instance, only ten episodes of *Radiolab*, a program with more than 500 episodes to date (Radiolab, 2023), were sampled in the initial study.

Another issue in the initial study is that the corpus used may not be fully representative of general-audience podcasts in terms of podcast types. A close reading of the transcripts of podcast programs suggested that the programs selected in the initial study can be roughly categorized into two types: talk shows and non-fiction narratives. In talk shows, a single host interviews one or more guests at a time by asking a set of questions, such as in *Fresh Air* and *How I Built This*. In non-fiction narratives, one or more hosts introduce a topic and invite guests to share personal stories and experiences in relation to that topic, such as in *Radiolab* and *This American Life*. However, news reports, as one of the major type of podcasts, were not included in the initial study. In news reports, one or more news anchors read news headlines or converse with other journalists.

In short, the aim of this study is to replicate Nurmukhamedov and Sharakhimov (2021) with a general-audience English podcast corpus that is much larger and more comprehensive than that of the initial study.

3. The replication study

The major variable changed in our replication was the corpus data. To be specific, we substantially increased the corpus size to provide a larger sample that is representative of general-audience English podcasts (see Table 1). To this end, we endeavored to increase the representativeness of our corpus (where feasible) in collecting the transcripts of different podcast programs.

Table 1. Podcast corpus in the initial study and the replication study

	Number of podcast programs	Number of episodes	Number of words	Duration (hours)
The initial study	9	170	1,137,163	112:33:05
The present study	12	8,862	14,498,043	1,460[a]

Note. [a]The podcast duration is an estimation of the number of hours.

First, while the initial study only examined two types of podcast programs (i.e., talk show and non-fiction narrative), the replication included three types (i.e., news report, talk show, and non-fiction narrative) to make the corpus data more comprehensive.

Second, in the initial study, balance was maintained across programs in terms of the number of episodes (i.e., 20 episodes for each podcast). In the replication, we kept the three sub-corpora (i.e., news report, talk show, and non-fiction narrative) balanced in terms of the number of podcast programs and overall number of running words. That is, we selected four programs for each sub-corpus, thus increasing the number of podcast programs from nine to 12. For each program, transcripts that were available were collected as much as possible without being limited by the number of episodes, until the overall numbers of words for the three sub-corpora were roughly balanced. As a result, 5.2 million words were sampled for news, 4.6 million for talk shows, and 4.7 million for non-fiction narratives. Finally, a 14-million-word corpus was compiled, and the number of podcast program episodes increased from 170 to 8,862.

Third, of the nine programs included in the initial study, we intentionally retained three programs (i.e., *Fresh Air*, *Radiolab*, and *This American Life*) in order to facilitate further comparisons. *Fresh Air* and *This American Life* were retained because they provided the largest number of episodes in their corresponding types of talk show and non-fiction narrative among the nine programs in the initial study. *Radiolab* was retained specifically due to the limited number of episodes sampled (i.e., ten episodes) in the initial study. All remaining variables from the initial study were kept unchanged.

The research questions that guided our study were exactly the same as the initial study. Specifically, the following two research questions were to be addressed:

RQ1. How many words do English language learners need to know to understand general-audience English podcasts?

RQ2. Will different podcasts programs draw on different vocabulary sizes to reach 95% and 98% coverage?

4. Methods

This section describes in detail the corpus data and analysis of lexical coverage.

4.1 Corpus data

To create the podcast corpus, the transcripts of 8,862 podcast episodes (boasting more than 14 million running words) were downloaded from the websites of 12 podcast programs (see Table 2 for statistics). The 12 podcast programs were chosen by following the three selection criteria of corpus data collection described in the initial study – that is, popularity, availability of transcripts, and a wide range of topics (Nurmukhamedov & Sharakhimov, 2021). First, all programs included in the replication are well-established in the podcast industry and have had a fairly long history of broadcasting. They were selected because all of them were in the top 100 podcast shows list according to a web-based radio service platform called Stitcher, and in the top 100 podcast shows in the iTunes charts (Nurmukhamedov & Sharakhimov, 2021). Second, in addition to the downloadable podcast audios,

Table 2. Description of the podcast corpus

Podcast types	Podcast title	Number of episodes	Number of words	Duration (hours)[a]
News				
	All Things Considered	2,000	1,467,552	147
	Morning Edition	2,982	1,956,909	194
	Weekend Edition	2,000	1,570,386	148
	Up First	106	271,182	29
Talk show				
	Death, Sex, and Money	284	1,236,146	129
	Fresh Air	250	867,734	95
	It's Been a Minute	169	980,559	104
	The Tim Ferriss Show	83	1,473,407	151
Non-fiction narrative				
	Snap Judgement	366	639,646	67
	Radiolab	191	1,415,912	116
	Rough Translation	65	328,304	34
	This American Life	366	2,290,306	246
Grand total		8,862	14,498,043	1,460

Note. [a]The podcast duration is an estimation of the number of hours.

the transcripts of these programs are also publicly available on the websites. More importantly, all transcripts were carefully checked by professionals before being released online. In other words, the availability of error-free transcripts made our data collection and analysis more feasible. Last, these programs covered a wide range of topics about daily life and can be considered as a snapshot of podcasts made for the general audience. Hence, the transcripts obtained from these programs can provide high-quality samples that are representative of general-purpose English podcast programs.

It may be of interest to note the similarities and differences between the corpus in the replication and that in the initial study. In terms of similarities, both corpora covered a wide range of topics, as each podcast program focuses on different topics in its episodes. For example, *Freakonomics Radio* in the initial study mainly discusses socioeconomic, political, educational, and psychological issues; *All Things Considered* selected in the replication study covers reports in arts and life, music, and entertainment; and *Radiolab*, selected in both studies, retells a series of science-based stories. In addition, the podcast programs included in both corpora have varied formats (i.e., the number of hosts and guests). Most of the selected programs have one host, and in each episode the host invites one or more guests to talk about an issue, such as *Hidden Brain* and *How I Built This* in the initial study, *The Tim Ferriss Show* and *Snap Judgement* in the replication study, and *Fresh Air* in both. There are also podcasts that have two or more hosts/anchors conversing with multiple guests/correspondents, such as *Invisibilia* in the initial study, *Morning Edition* in the replication study, and *Radiolab* in both.

Some notable differences also exist between the two corpora. The first major difference is the number of episodes. While the initial study collected 20 episodes for each podcast program, our replication had a much wider episode range of 65 to 2,982 across programs. The reason for this difference lies in the fact that the initial study kept the balance in terms of episode numbers across podcast programs, whereas our replication study maintained the balance in terms of the number of

words for each podcast type. That said, our corpus would have been limited by the lowest number of episodes available among the 12 programs (65 episodes in this case) if an episode-number-based sampling strategy was adopted. Second, due to the inclusion of the news report category, which is usually shorter in length than that of talk show and non-fiction narrative, the podcasts selected in the replication study are notably shorter on average than those of the initial study. Specifically, the average number of words per episode in the replication was 1,636, ranging from 656 (i.e., *Morning Edition*) to 17,752 (i.e., *The Tim Ferriss Show*), while that in the initial study was 6,689, ranging from 1,414 (i.e., *Radiolab*) to 10,601 (i.e., *This American Life*). The duration per episode in the replication was 10 min, with the shortest program lasting for about 4 min per episode (i.e., *Morning Edition*) and the longest one 109 min (i.e., *The Tim Ferriss Show*). In comparison, the average duration per episode was 40 min in the initial study, ranging from 7 min (i.e., *Radiolab*) to 62 min (i.e., *This American Life*).

4.2 Data coding and analysis

We followed the same data analysis procedures described in Nurmukhamedov and Sharakhimov (2021). As in the initial study, to determine the vocabulary size needed to reach 95% and 98% lexical coverage, the analysis of podcast transcripts was performed using Laurence Anthony's vocabulary profiling tool, *AntWordProfiler* (Anthony, 2023), loaded with the BNC/COCA word family lists (25 1,000-word lists). Note that four additional lists – that is, proper nouns (PN), marginal words (MW), transparent compounds (TC), and acronyms (AC) – were also included in the BNC/COCA word family lists. Words that were not matched in the foregoing lists were categorized as *Not in the lists*.

A preliminary analysis was carried out for our corpus by using *AntWordProfiler* (Anthony, 2023). Then, similar to the initial study, the following modifications were made to ensure that the analysis of lexical coverage in the podcast transcripts was reliable. First, contractions, connected speech, and hyphenated words were changed to conform with the spelling scheme implemented in the BNC/COCA word family lists. In the initial study, contractions (e.g., *she's* and *we've*) and connected speech (e.g., *wanna* and *kinda*) were changed into their full form (e.g., *she is*, *we have*, *want to*, and *kind of*). Hyphens in compound words were replaced by spaces so that the two words comprising the compound would be classified according to their respective frequency in the BNC/COCA word lists. Given the size of the corpus, a home-made python script was coded to change the spellings of the aforementioned word categories, ensuring they would not be classified as *Not in the lists* words. Second, proper nouns and acronyms that were used in the transcripts but were not correctly classified as PN and AC in the analysis were manually reclassified and added to the original PN and AC lists. For instance, words like *Messi* and *TikTok* were reclassified as PN and added to the PN lists; words like *COVID-19* and *USOPC* were reclassified as AC and added to the AC lists. Last, company names (e.g., *ByteDance* and *Uber*), social networking services (e.g., *WeChat* and *Facebook*), locations (e.g., *Ashville* and *Shantou*), and ethnic names (e.g., *Schwartzel* and *Falkowski*) were reclassified and added to the PN list. To ensure the reliability of the manual reclassification, the two researchers worked together and resolved all disputed cases through discussions.

The corpus was reanalyzed by *AntWordProfiler* (Anthony, 2023) using the modified BNC/COCA word lists. *AntWordProfiler* allowed us to know the distribution of words at different frequency levels, and know the number of word families required to reach 95% and 98% coverage either with or without the PN, WM, TC, and AC. The results are presented in Table 3.

5. Results and discussion

This section reports and discusses in detail the major findings of the replication study.

5.1 Research Question 1

Table 3 presents the number of word families and their proportion at each frequency level in the BNC/COCA word list. Results of the lexical coverage analysis suggested that knowledge of the 3,000 most

Table 3. Cumulative lexical coverage in percentage of all podcasts

Word family	Coverage without PN, WM, TC, and AC	Coverage with PN, WM, TC, and AC
1,000	83.72	88.42
2,000	88.97	93.67
3,000	91.91	96.61[a]
4,000	92.89	97.59
5,000	93.47	98.17[b]
6,000	93.88	98.58
7,000	94.13	98.83
8,000	94.33	99.03
9,000	94.47	99.17
10,000	94.57	99.27
11,000	94.66	99.36
12,000	94.72	99.42
13,000	94.78	99.48
14,000	94.82	99.52
15,000	94.85	99.55
16,000	94.87	99.57
17,000	94.88	99.58
18,000	94.91	99.61
19,000	94.92	99.62
20,000	94.93	99.63
21,000	94.94	99.64
22,000	94.95	99.65
23,000	94.96	99.66
24,000	94.96	99.66
25,000	94.97	99.67
Proper nouns (PN)	3.64	
Marginal words (MW)	0.6	
Transparent compounds (TC)	0.36	
Acronyms (AC)	0.1	
Not in the lists	0.33	

Note. [a]Denotes reaching 95% coverage. [b]Denotes reaching 98% coverage.

frequent word families plus PN, WM, TC, and AC provided 96.61% coverage, while knowledge of the 5,000 most frequent word families plus the PN, WM, TC, and AC provided 98.17% coverage. The results were consistent with findings from the initial study, which found that podcast listeners needed a vocabulary of 3,000 and 5,000 word families plus the knowledge of PN, WM, TC, and, AC to gain 95% and 98% coverage, respectively. The coverage percentages in our replication (i.e., 96.61% and 98.17%) were slightly different from those in the initial study (i.e., 96.75% and 98.26%). If 95% coverage is deemed sufficient for comprehension, the 3,000 most frequent word families can be considered as an attainable goal in EFL settings. Laufer (2001), for example, found that Chinese English majors

Table 4. Comparison of word coverage in the initial study and the replication study

Category	The initial study	The replication study
High-frequency words	91.17%	91.91%
Mid-frequency words	2.43%	2.56%
Low-frequency words	0.57%	0.50%
PN, WM, TC, and AC	5.58%	4.60%
Not in the lists	0.26%	0.33%

had a vocabulary size of 4,000 word families. Similarly, Ozturk (2012) found that advanced EFL leaners in a Turkish university had a vocabulary size of 3,200–7,900 word families. Even if 98% is established for adequate comprehension, 2,000 more word families are still attainable with additional assistance. This confirms that general-audience English podcasts can serve as an appropriate source of L2 input material in terms of lexical demand.

A closer look into the cumulative lexical coverage for the running words showed that the lexical profile of general-audience English podcasts in the replication study was comparable with that in the initial study. As is shown in Table 4, high-frequency words provided similar coverage in the corpus of the initial study and our corpus (i.e., 91.17% vs. 91.91%), as did mid-frequency words (i.e., 2.43% vs. 2.56%) and low-frequency words (i.e., 0.57% vs. 0.50%). The four supplementary lists of PN, WM, TC, and AC accounted for 5.58% of running words in the initial study and 4.60% in our replication, second only to high-frequency words. This point highlights the significance of including PN, WM, TC, and AC in profiling lexical coverage of general-audience podcasts. That said, if PN, WM, TC, and AC were not assumed to be known, even 25,000 words families were insufficient to gain 95% coverage, which also corroborated findings from the initial study.

It may also be interesting to compare the lexical demand of general-audience English podcasts to that of different types of spoken discourse so that teachers can assess their suitability as authentic listening materials for EFL teaching. Overall, general-audience English podcasts are less demanding than spoken discourse involving academic content, such as TED talks (Nurmukhamedov, 2017), academic podcasts (Liu, 2023), and university-based academic lectures (Dang & Webb, 2014). In addition, general-audience podcasts are similar to most authentic scripted and unscripted spoken discourse for general purpose in terms of lexical demand. They are comparable to movies (Webb & Rodgers, 2009a), television programs (Webb & Rodgers, 2009b), charted songs and teacher-selected songs (Tegge, 2017), soap operas and sitcoms (Al-Surmi, 2014), the listening section of Test of English as a Foreign Language (TOEFL) internet-based test (Kaneko, 2015), and university tutorials and laboratory sessions (Coxhead et al., 2017). However, in comparison with recent findings on English-as-an-additional-language (EAL) podcasts (Motamedynia & Shahri, 2022), general-audience English podcasts are more demanding in lexical coverage (see Appendix 1 for details).

Taken together, when 95% and 98% coverage levels are examined, general-audience English podcasts are located somewhere in the middle of the lexical demand continuum (Motamedynia & Shahri, 2022). More precisely, they are located somewhere towards the lower side.

5.2 Research Question 2

The second question relates to the variation of the vocabulary demands in different podcast programs. Overall, vocabulary demands necessary for 95% coverage (i.e., 3,000 word families) were fairly consistent among most podcast programs (nine out of 12) (see Table 5). Interestingly, three programs (i.e., *Death, Sex, and Money*, *It's Been a Minute*, and *Snap Judgement*) required only 2,000 word families plus PN, WM, TC, and AC for adequate comprehension. In contrast, with reference to 98% coverage, vocabulary demands varied from 4,000–6,000 word families plus PN, WM, TC and, AC. To be

Table 5. Cumulative lexical coverage in percentage by podcast

Podcast program	1,000*	2,000*	3,000*	4,000*	5,000*	6,000*
All Things Considered	85.1	91.92	96.21[a]	97.42	98.1[b]	98.62
Morning Edition	85.27	92.05	96.39[a]	97.55	98.2[b]	98.72
Weekend Edition	86.79	92.72	96.28[a]	97.44	98.09[b]	98.56
Up First	85.35	92.06	96.31[a]	97.33	98.02[b]	98.62
Death, Sex, and Money	92.17	95.82[a]	97.48	98.1[b]	98.47	98.76
Fresh Air	88.16	93.36	96.24[a]	97.27	97.84	98.26[b]
It's Been a Minute	90.95	95.02[a]	97.17	97.88	98.31[b]	98.68
The Tim Ferriss Show	88.72	93.75	96.68[a]	97.58	98.09[b]	98.42
Snap Judgement	90.84	95.12[a]	96.7	97.62	98.26[b]	98.65
Radiolab	89.54	94.14	96.41[a]	97.34	97.96	98.35[b]
Rough Translation	89.6	94.19	96.78[a]	97.64	98.18[b]	98.51
This American Life	89.98	94.56	96.79[a]	97.74	98.32[b]	98.71

Note. *The cumulative percentage includes proper nouns, marginal words, transparent compounds, and acronyms. [a]Denotes reaching 95% coverage. [b]Denotes reaching 98% coverage.

specific, *Death, Sex, and Money* was the least demanding program, requiring 4,000 word families plus PN, WM, TC, and AC for 98% coverage, whereas *Fresh Air* and *Radiolab* were the most demanding, requiring 6,000 word families plus PN, WM, TC, and AC for 98% coverage. Both the vocabulary demands necessary for 95% and 98% coverage were similar to those of the initial study, hence lending full support to the findings of the initial study.

In our replication, we intentionally retained three programs from the initial study – that is, *Fresh Air*, *Radiolab*, and *This American Life*. On the one hand, the coverage figures in our study and the initial study revealed that *Radiolab* and *This American Life* exhibited similar vocabulary demands to reach 95% and 98%. That is, *Radiolab* needed 3,000–6,000 word families plus PN, WM, TC, and AC, while *This American Life* required 3,000–5,000 plus PN, WM, TC, and AC. On the other hand, for *Fresh Air*, knowledge of 3,000–5,000 word families plus PN, WM, TC, and AC was required in the initial study, but knowledge of 3,000–6,000 word families plus PN, WM, TC, and AC was needed in our replication. Note that the 1,000-word family difference matters because native speakers may learn approximately 1,000 word families per year (Goulden et al., 1990), while L2 speakers may only learn 400–500 words per year (Ozturk, 2012; Webb & Chang, 2012).

What may lead to the 1,000-word family difference between the two studies? Nurmukhamedov and Sharakhimov (2021) argued that three factors might affect the vocabulary size figures of podcasts – that is, formats (i.e., the number of hosts and guests), topics, and disciplines. Based on further examination of the *Fresh Air Archive* (available at https://freshairarchive.org), we found that all episodes in *Fresh Air* had the same format, featuring long-form interviews conducted by the same host. Thus, the difference in vocabulary demand might be pertinent to different topics and disciplines. Aired since 1985, *Fresh Air* has 868 topic tags falling into 22 categories (as of 15 September 2023), such as Business and Economy, Art, and Science and Technology. As the initial study only sampled 20 episodes while our replication sampled 250, the inclusion of more topics and disciplines were likely to cause an increase of 1,000 word families in lexical demand, particularly when the sampled transcripts included topics in Science and Technology, such as physics, neuroscience, and epidemiology (Dang & Webb, 2014). This suggests the importance of sample size for better representativeness of individual programs and more stable size figures in lexical demand analysis (see Table 6).

Of the three retained podcasts, results of our replication indicated that the proportion of PN, WM, TC, and AC in *Radiolab* was consistent in the present study and the initial study (see Table 7). That is,

Table 6. Comparison of word coverage in the initial study and the replication study

Podcast program	Study	Number of episodes	Number of words	Duration (hours)
Fresh Air	Replication	250	867,734	95[a]
	Initial	20	131,057	11:53:28
Radiolab	Replication	191	1,415,912	116[a]
	Initial	10	14,137	01:12:58
This American Life	Replication	366	2,290,306	246[a]
	Initial	20	212,020	20:37:08

Note. [a]The podcast duration is an estimation of the number of hours.

Radiolab had the highest percentage of PN, WM, TC, and AC in both studies: 9.69% in the initial study and 7.73% in the replication study. However, we also noticed some differences, particularly a marked discrepancy in *This American Life*. In the initial study, the knowledge of PN, WM, TC, and AC accounted for 7.7%, the second highest percentage among nine podcasts, while the number was 3.71% in our study. To investigate the reason underlying this discrepancy, we logged onto its website (www.thisamericanlife.org) and found that this program had 809 aired episodes (as of 15 September 2023). Given that the initial study only collected 20 episodes while our study sampled 366 episodes, this again indicated that the sample size of corpus data had an effect on the vocabulary size analysis.

6. Summary of the replication

We conducted a close replication of Nurmukhamedov and Sharakhimov's (2021) research, which examined the lexical profile of general-audience English podcasts. To this end, we collected our corpus data following the selection criteria provided in the initial study, but with modifications in corpus compilation. First, the corpus size was substantially increased to provide better representative samples of the general-audience English podcasts. The initial study compiled a 1-million-word corpus, comprising 112 running hours of podcasting, while our replication used a 14-million-word corpus, totaling

Table 7. Percentage of PN, WM, TC, and AC in the replication study and the initial study

Podcast program	Percentage of PN, WM, TC, and AC in the replication	Percentage of PN, WM, TC, and AC in the initial study
All Things Considered	5.72	–
Morning Edition	5.96	–
Weekend Edition	5.46	–
Up First	5.82	–
Death, Sex, and Money	3.65	–
Fresh Air	3.77	4.4
It's Been a Minute	3.79	–
The Tim Ferriss Show	2.23	–
Snap Judgement	3.24	–
Radiolab	7.73	9.69
Rough Translation	5.13	–
This American Life	3.71	7.7

1,460 running hours. Second, the initial study included 170 episodes of podcast programs in total. Our replication significantly increased the number of selected episodes to 8,862 to better represent general-audience English podcast programs. Last, the initial study included nine general-audience podcast programs. Our replication intentionally retained three programs included in the initial study and added another nine, making up a total of 12 podcast programs. It should also be noted that the initial study only examined two types of podcast programs (i.e., talk show and non-fiction narrative). Our replication included three types (i.e., news report, talk show, and non-fiction narrative) to make the corpus data more comprehensive. The remaining aspects of the initial study were kept the same in the replication.

Overall, our results corroborated the initial study in major findings. The findings that mirrored the initial study included: (1) knowledge of 3,000 and 5,000 word families plus the knowledge of PN, WM, TC, and AC were needed to gain 95% and 98% coverage, respectively; (2) vocabulary demands necessary for 95% coverage were fairly consistent among most podcast programs, while vocabulary demands necessary for 98% coverage varied from 4,000–6,000 word families; and (3) of the three retained podcasts, *Radiolab* and *This American Life* exhibited similar vocabulary demands to reach 95% and 98%. Hence, findings from this replication study provided supporting evidence of the vocabulary demand of general-audience English podcast programs. However, we also noticed some minor differences in some individual podcasts. These differences included: (1) vocabulary demands to reach the 98% coverage in *Fresh Air* were different in the initial study and our replication (5,000 vs. 6,000 plus PN, WM, TC, and AC); and (2) discrepancy occurred in *This American Life* in terms of the percentage of PN, WM, TC, and AC in the initial study and our replication (7.73% vs. 3.71%). These differences may be pertinent to differences in data sampling, as procedures in corpus analysis were kept unchanged.

Our findings have important implications for lexical coverage research. The first implication is the importance of using a sample size as large as possible in vocabulary coverage analysis to better represent the target discourse. As formats, topics, and disciplines are potential factors that may affect the coverage percentage of podcasts (Nurmukhamedov & Sharakhimov, 2021), the inclusion of more data could contribute to the stability of the lexical coverage figures. Moreover, when more sample episodes are included, individual podcast programs are better represented, which provides a suitable basis for researchers to make more valid and nuanced observations (Ädel, 2020; McEnery & Hardie, 2012), thus increasing the reliability and generalizability of the lexical coverage figures. The second implication is the necessity of conducting replication studies in the area of lexical coverage. Using the same methodology but testing in a much larger and comprehensive corpus, this replication study offers insights into the extent to which the intentional one-variable modification might shape the conclusions. The similarities between the two studies provide evidence in support of the initial study's findings regarding the lexical demand and lexical difficulty of general-audience English podcasts in relation to other types of spoken discourse. The differences indicate the possibility of fine-grained investigations and comparative analysis that otherwise would be neglected without the replication study. In this respect, both similarities and differences observed in the present study contributed to "a more reliable, nuanced, and ecologically valid understanding" (Schmitt et al., 2017, p. 214) of the lexical profile of general-audience English podcasts.

The findings also have important implications for EFL teaching and learning. To start with, general-audience English podcasts cannot be simply treated as entry-level listening materials, although they are located somewhere towards the lower end in the lexical demand continuum for spoken discourse. There are several reasons for this point. First, the high percentage of PN, WM, TC, and AC, particularly PN, may pose a great challenge for comprehension. In lexical coverage research, it is a standard practice to assume that PN are unproblematic for L2 learners (Nation, 2006; Webb & Rodgers, 2009a, 2009b) because "proper nouns are not lexical items" (Cobb, 2010, p. 187). However, previous studies suggested that unfamiliar PN could interrupt the flow of reading (Brown, 2010) and listening (Kobeleva, 2008), thus placing some learning burdens on L2 learners (Kobeleva, 2008). Second, the length of general-audience podcasts can be considered as another factor that may affect comprehension. Of the 12 programs selected for replication, more than half of them

had more than 3,000 running words per episode on average. For comparison, charted songs had an average of 435 running words (Tegge, 2017). Although the lexical demand of charted songs seemed to be higher (i.e., 3,000 and 6,000 plus PN, TC, and MW to reach 95% and 98% coverage) (Tegge, 2017), songs can be more advantageous as entry-level listening and reading-for-listening materials due to the brevity of lyrics compared to general-audience podcasts. Third, general-audience podcasts are authentically-sourced listening materials, which may be difficult for L2 learners to handle. As general-audience podcasts are "definitely not created with language learners in mind" (Nurmukhamedov & Sadler, 2011, p. 182), L2 learners who are suddenly thrust into a native speaker environment may be stunned at the rate of speech and confused by the native-like pronunciation, such as connected speech. Therefore, when selecting general-audience podcasts either for pedagogical purposes or material development, teachers need to be aware that the use of general-audience podcasts with lower-level students could be problematic. Hence, it is more proper to consider general-audience podcasts as mid-level materials (Motamedynia & Shahri, 2022).

As L2 learners, particularly lower-level students, may encounter comprehension problems in listening, teachers should employ strategies to ease their burden on comprehension. One strategy they can employ is differentiated instruction. For example, teachers can choose a less-lexically-demanding podcast, such as *Death, Sex, and Money* (with a lexical demand of 2,000–4,000 word families) for lower-level students for extra-curriculum listening. Meanwhile, they can select a slightly more demanding podcast, such as *It's Been a Minute* or *Snap Judgement* (with a lexical demand of 2,000–5,000 word families) for intermediate-level students, and a podcast like *This American Life* (with a lexical demand of 3,000–5,000 word families) for advanced students. In cases where students might need additional assistance, teachers can provide "pre-teaching key or low-frequency vocabulary essential for comprehension" (Nurmukhamedov & Sharakhimov, 2021, p. 11), or transcripts of podcast programs to support listening. Students can "rewind, forward, or pause the text" (Liu, 2023, p. 19) or "use a slow-down feature in their smart phones" (Nurmukhamedov & Sharakhimov, 2021, p. 13) based on their own needs to facilitate comprehension. In such cases, L2 learners may maximize their comprehension by using a level-appropriate podcast and receiving a necessary amount of assistance.

When assigning or recommending podcasts to L2 learners, teachers should be aware that findings pertaining to the whole podcast genre are not necessarily applicable to individual programs. Similarly, findings pertaining to individual programs are not necessarily applicable to each episode. Neither the initial study nor our replication study carried out an exploratory analysis of lexical demands of the individual episodes in a program. However, prior research has shown that the distribution of lexical demands among different episodes might vary greatly (Liu, 2023). Therefore, L2 teachers are encouraged to investigate the vocabulary demand of an individual episode before it is assigned or recommended to students. A randomly selected episode may result in poor comprehension and demotivate the students if it is lexically too easy or too demanding.

While vocabulary load is of concern to teachers when selecting general-audience podcasts for pedagogical purposes, they should also pay attention to other characteristics of podcasts, such as length. As noted, podcast length can vary greatly, from as low as 656 running words per episode on average (e.g., *Morning Edition*), to as high as 17,752 running words per episode (e.g., *The Tim Ferriss Show*), according to the sampling in our replication. The length of transcripts can determine for what purpose and how each episode can be used. For example, short episodes in *Morning Edition* can be used for intensive listening or reading and intentional vocabulary learning in a language-learning class, as their brevity makes it possible to be used in their entirety and for repeated listening or reading. On the other hand, long episodes in *The Tim Ferriss Show* can be used as extra-curriculum materials for extensive listening or reading and incidental vocabulary learning.

7. Implications for future study

Although similar results have been obtained in the replication study, additional work is still needed in the future. First, future studies may use a different counting unit. While the word family unit is widely

used, a smaller unit (e.g., flemma or lemma) may be the more suitable counting unit for L2 vocabulary research and pedagogy (Brown et al., 2020, 2021; McLean, 2018). Word-family-based research assumes that learners with the knowledge of a base word can comprehend other members (i.e., inflections and derivations of the base word) (Nation, 2006), but empirical findings suggested that the knowledge of a base word often fail to correspond with the knowledge of its family members, particularly for L2 learners (Kremmel & Schmitt, 2016; McLean, 2018; Ward & Chuenjundaeng, 2009). Lemma (i.e., a baseword of a particular part of speech and inflections) and flemma (i.e., a base form and inflectional forms, regardless of part of speech), on the other hand, are lexical units that involve less learning burdens for L2 learners (Schmitt, 2010) and align better with their abilities. Thus, it would be interesting to observe whether a different choice of counting unit makes a difference in lexical demand estimates and whether word-family-based studies underestimate the amount of words needed.

In addition, the choice of different counting units involves the use of different word lists. Schmitt et al. (2017) proposed the use of Mark Davies' lemmatized frequency list of the complete COCA (available at https://www.wordfrequency.info/intro.asp). Future studies using Davies' lemma word lists can compare them with studies using Nation's word family word lists, shedding light on "how generalizable Nation's word family figures are for pedagogical purposes" (Schmitt et al., 2017, p. 219).

Last, future studies may re-examine the domains that were previously studied, particularly those using the original BNC-based frequency lists (Al-Surmi, 2014; Dang & Webb, 2014; Nurmukhamedov, 2017; Tegge, 2017). Since the updated BNC/COCA-based lists are a better indication of the frequency information, revisiting these domains with the new lists may produce quite different results.

References

Abraham, L. B., & Williams, L. (Eds.) (2009). *Electronic discourse in language learning and language teaching*. John Benjamins.

Ädel, A. (2020). Corpus compilation. In M. Paquot & S. T. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 3–24). Springer.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425–438. doi:10.1093/applin/24.4.425

Al-Surmi, M. (2014). TV shows, word coverage and incidental vocabulary learning: Implications for Arabic speakers learning English. In K. M. Bailey & R. M. Damerow (Eds.), *Teaching and learning English in the Arabic-speaking world* (pp. 132–147). Routledge.

Anthony, L. (2023). *Antwordprofiler (Version 4.2.4)*. Waseda University. Retrieved from https://www.laurenceanthony.net/software

Brown, D. (2010). An improper assumption? The treatment of proper nouns in text coverage counts. *Reading in a Foreign Language*, *22*(2), 355–361. http://hdl.handle.net/10125/66842

Brown, D., Stoeckel, T., McLean, S., & Stewart, J. (2020). The most appropriate lexical unit for L2 vocabulary research and pedagogy: A brief review of the evidence. *Applied Linguistics*, *43*(3), 596–602. doi:10.1093/applin/amaa061

Brown, D., Stewart, J., Stoeckel, T., & McLean, S. (2021). The coming paradigm shift in the use of lexical units. *Studies in Second Language Acquisition*, *43*(5), 950–953. doi:10.1017/S0272263121000668

Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language*, *22*(1), 181–200. http://hdl.handle.net/10125/66641

Coxhead, A., Dang, T. N. Y., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, *30*(4), 66–78. doi:10.1016/j.jeap.2017.11.001

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*(2), 66–76. doi:10.1016/j.esp.2013.08.001

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology*, *13*(3), 66–86. http://hdl.handle.net/10125/44192

Egbert, J., & Baker, P. (Eds.) (2016). *Triangulating methodological approaches in corpus linguistic research*. Routledge.

Facer, B. R., & Abdous, M. (Eds.) (2011). *Academic podcasting and mobile assisted language learning: Applications and outcomes*. Information Science Reference.

Facer, B. R., Abdous, M., & Camarena, M. M. (2009). The impact of podcasting on students' learning outcomes. In R. de. C. V. Marriott & P. L. Torres (Eds.), *Handbook of research on e-learning methodologies for language acquisition* (pp. 339–351). Information Science Reference.

Fouz-González, J. (2019). Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features. *ReCALL*, *31*(2), 150–169. doi:10.1017/S0958344018000174

Goetzen, N. (2020). *Time spent with podcasts will dip amid the pandemic, but it should rebound by 2022*. eMarketer. Retrieved from https://www.insiderintelligence.com/content/time-spent-with-podcasts-will-dip-amid-pandemic-should-rebound-by-2022

Goulden, R., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363. doi:10.1093/applin/11.4.341

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and Reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430. http://hdl.handle.net/10125/66973

Kaneko, M. (2015). Vocabulary size required for the TOEFL iBT listening section. *The Language Teacher*, *39*(1), 9–14. doi:10.37546/JALTTLT39.1-2

Kobeleva, P. P. (2008). The impact of unfamiliar proper names on ESL learners' listening comprehension [Unpublished Ph.D. Dissertation]. Victoria University of Wellington.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*(4), 377–392. doi:10.1080/15434303.2016.1237516

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. (2001). Quantitative evaluation of vocabulary: How it can be done and what it is good for. In C. Elder & A. Davies (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 241–250). Cambridge University Press.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size. *Reading in a Foreign Language*, *22*(1), 15–30. http://hdl.handle.net/10125/66648

Liu, C.-Y. (2023). Podcasts as a resource for learning academic English: A lexical perspective. *English for Specific Purposes*, *71*(3), 19–33. doi: 10.1016/j.esp.2023.02.003

McBride, K. (2009). Podcasts and second language learning: Promoting listening comprehension and intercultural competence. In L. B. Abraham & L. Williams (Eds.), *Electronic discourse in language learning and language teaching* (pp. 153–168). John Benjamins.

McEnery, T., & Brookes, G. (2022). Building a written corpus: What are the basics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 35–47). Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, *39*(6), 823–845. doi:10.1093/applin/amw050

McManus, K. (2022). Are replication studies infrequent because of negative attitudes? Insights from a survey of attitudes and practices in second language research. *Studies in Second Language Acquisition*, *44*(5), 1410–1423. doi:10.1017/S0272263121000838

Motamedynia, M., & Shahri, N. N. (2022). Investigating the lexical demands of English-as-an-additional-language and general-audience podcasts and their potential for incidental vocabulary learning. *Canadian Journal of Applied Linguistics*, *25*(2), 103–131. doi:10.37213/cjal.2022.32746

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59–82. doi:10.1353/cml.2006.0049

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

Nation, I. S. P. (2018). *The BNC/COCA word family lists*. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P. (2020). *The BNC/COCA word family lists*. Retrieved from https://www.wgtn.ac.nz/__data/assets/pdf_file/0005/1857641/about-bnc-coca-vocabulary-list.pdf

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Nurmukhamedov, U. (2017). Lexical coverage of TED talks: Implications for vocabulary instruction. *TESOL Journal*, *8*(4), 768–790. doi:10.1002/tesj.323

Nurmukhamedov, U., & Sadler, R. (2011). Podcasts in four categories: Applications to language learning. In B. R. Facer & M. Abdous (Eds.), *Academic podcasting and mobile assisted language learning: Applications and outcomes* (pp. 176–195). Information Science Reference.

Nurmukhamedov, U., & Sharakhimov, S. (2021). Corpus-based vocabulary analysis of English podcasts. *RELC Journal*, *54*(1), 7–21. doi:10.1177/0033688220979315

Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, *52*(2), 188–200. doi:10.1017/S0261444819000028

Omidian, T., Ballance, O. J., & Siyanova-Chanturia, A. (2023). Replicating corpus-based research in English for academic purposes: Proposed replication of Cortes (2013) and Biber and Gray (2010). *Language Teaching*, *56*(1), 128–136. doi:10.1017/S0261444821000367

Ozturk, M. (2012). Second language vocabulary growth at advanced level. *The Language Learning Journal*, *44*(1), 6–16. doi:10.1080/09571736.2012.708054

Porte, G. K. (2012). *Replication research in applied linguistics*. Cambridge University Press.

Porte, G. K., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Quah, N. (2021). *Yes, podcast listenership is still on the rise*. Vulture. Retrieved from https://www.vulture.com/2021/03/podcast-listenership-download-data-on-the-rise.html

Radiolab. (2023). *Podcast*. Retrieved from https://www.radiolab.org/podcast

Reppen, R. (2022). Building a corpus: What are key considerations? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 13–20). Routledge.

Rodgers, M. P. H., & Webb, S. (2016). Listening to lectures. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 165–176). Routledge.

Rowe, A. (2020). *How the podcast industry is surviving the pandemic*. Tech.co. Retrieved from https://tech.co/news/podcasting-pandemic-impact-2020-07

Saeedakhtar, A., Haqju, R., & Rouhi, A. (2021). The impact of collaborative listening to podcasts on high school learners' listening comprehension and vocabulary learning. *System*, *101*(1), 1–15. doi:10.1016/j.system.2021.102588

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

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

Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, *50*(2), 212–226. doi:10.1017/S0261444815000075

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System*, *67*(4), 87–98. doi:10.1016/j.system.2017.04.016

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. doi:10.1093/applin/ams074

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, *37*(3), 461–469. doi:10.1016/j.system.2009.01.004

Webb, S. (2021). Research investigating lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language*, *33*(2), 278–293. http://hdl.handle.net/10125/67407

Webb, S., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*(1), 113–126. doi:10.1177/0033688212439367

Webb, S., & Paribakht, T. S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes*, *38*(1), 34–43. doi:10.1016/j.esp.2014.11.001

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427. doi:10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366. doi:10.1111/j.1467-9922.2009.00509.x

Whitner, G. (2023). *The meteoric rise of podcasting: Insights about the most compelling audio format*. Music Strive. Retrieved from https://musicstrive.com/podcast-statistics/

Appendix 1. Lexical coverage in spoken discourse

Study	Lexical coverage topics	Coverage figure (%) and number of word families
Adolphs and Schmitt (2003)	General spoken English	(95) 3,000 (96) 5,000
Nation (2006)	General spoken English	(98) 6,000–7,000
Webb and Rodgers (2009a)	Movies	(95.76) 3,000 (98.15) 6,000
Webb and Rodgers (2009b)	Television programs	(95.45) 3,000 (98.27) 7,000
Dang and Webb (2014)	Academic spoken English	(96.05) 4,000 (98.00) 8,000
Al-Surmi (2014)	Soap opera /Sitcom	Soap opera (95.49) 2,000 (98.19) 5,000 Sitcom (95.06) 2,000 (98.07) 7,000
Webb and Paribakht (2015)	Listening section of CanTEST	(95.39) 4,000 (98.04) 10,000
Kaneko (2015)	Listening section of TOEFL internet-based test	(95) 3,000 (98.18) 6,000
Tegge (2017)	Pop songs	Charts corpus (95.87) 3,000 (98.16) 6,000 Pedagogical corpus (95.51) 2,000 (98.17) 5,000
Nurmukhamedov (2017)	TED talks	(95.89) 4,000 (98.07) 8,000
Coxhead et al. (2017)	University laboratory sessions and tutorials	Laboratories (96) 3,000 (98.0) 7,000 Tutorials (97.77) 3,000 (98.45) 4,000
Nurmukhamedov and Sharakhimov (2021)	General-audience podcasts	(96.75) 3,000 (98.26) 5,000
Motamedynia and Shahri (2022)	English-as-an-additional-language (EAL) podcasts	(95.42) 2,000 (98.14) 4,000
Liu (2023)	Academic podcasts	(95.78) 4,000 (98.04) 10,000

Hong Yu is Associate Professor of English at Southwest Petroleum University, China. Her research interests include corpus linguistics, translation studies, and EFL writing. Her most recent reviews have been published by *Journal of Second Language Writing*, *System*, and *RELC Journal*.

Ju Wen is Associate Professor of Applied Linguistics at Chengdu Jincheng College, China. He is interested in corpus linguistics, L2 writing, language education, and science communication. He has published in journals such as *Applied Linguistics*, *Journal of Second Language Writing*, *Public Understanding of Science*, *Journal of Education for Teaching*, and *Scientometrics*.