

# Exploring the Vocabulary Makeup of Scripted and Unscripted Television Programs and Their Potential for Incidental Vocabulary Learning

*Hesamoddin Shahriari and Masoud Motamedynia*

---

*The present study investigated the lexical demands of scripted and unscripted television programs. To that end, two corpora consisting of 286 episodes from 14 different programs, both scripted and unscripted, were analyzed. The results indicated that the 1,000 most frequent word families, plus proper nouns, marginal words, transparent compounds, and acronyms, were required to reach 90% coverage in both scripted and unscripted programs. Furthermore, knowledge of the 2,000 most frequent word families accounted for 95% coverage in the unscripted programs, while, to reach the same threshold in the scripted programs, a vocabulary size of the 3,000 most frequent word families was needed. Regarding 98% coverage, vocabulary knowledge of 4,000 and 6,000 word families was required for the unscripted and scripted programs, respectively. A corpus-driven investigation was also conducted to explore the potential of both types of television programs for incidental vocabulary learning. Accordingly, the results showed that both types of programs may hold relatively great potential for learning words from the 2,000- to 3,000-word levels and might have some potential for the incidental learning of mid-frequency words (i.e., 4,000- to 9,000-word levels). Implications for using both types of television programs in language learning and teaching processes are discussed.*

*La présente étude a examiné les niveaux lexicaux de séries télévisées présentant de l'oral préparé et de l'oral spontané. Afin de ce faire, deux corpus composés de 286 épisodes de 14 séries télévisées différentes, présentant de l'oral spontané et de l'oral préparé, ont été analysés. Les résultats indiquent que la connaissance des 1000 familles de mots les plus fréquentes ainsi que les noms propres, les interjections, les mots composés transparents et les acronymes était requise pour atteindre une couverture de 90 % pour les séries en oral spontané et en oral préparé. De plus, la connaissance des 2000 familles de mots les plus fréquentes représentait une couverture de 95 % pour les séries en oral spontané tandis qu'une connaissance des 3000 familles de mots les plus fréquentes était nécessaire pour atteindre ce même seuil pour les séries en oral préparé. Quant à*

*une couverture de 98 %, une connaissance lexicale des 4000 et 6000 familles de mots était requise pour les séries en oral spontané et les séries en oral préparé, respectivement. Une analyse axée sur le corpus a également été menée dans le but d'explorer le potentiel des deux types de séries quant à l'apprentissage incident du vocabulaire. Ainsi, les résultats démontrent que les deux types de séries possèderaient un potentiel relativement élevé de l'apprentissage des mots des 2000 à 3000 familles de mots et pourraient posséder un certain potentiel de l'apprentissage incident des mots à fréquence moyenne (à savoir les 4000 à 9000 familles de mots). Les implications de l'usage des deux types de séries télévisées dans l'apprentissage et l'enseignement des langues sont abordées.*

---

*Keywords:* incidental vocabulary learning, lexical coverage, scripted, television, unscripted

---

Television plays an important role in providing first language input in large quantities. A survey conducted by the Organization of Economic Co-operation and Development shows that people around the world spend between two (e.g., Norway) and eight hours (e.g., the United States) per day watching television in their native language (OECD, 2009). There is also no doubt that television is a potentially valuable source of aural input for second and foreign language learning (e.g., Lindgren & Muñoz, 2013; Peters, 2018; Rodgers & Webb, 2020). Thanks to the advances in information technology and high-speed internet, EFL learners now have easier access to linguistic audiovisual input, including television programs, in English.

Adequate knowledge of L2 vocabulary is one important factor influencing learners' comprehension and, as a result, their television viewing habits (Webb, 2015). Such knowledge has been suggested to play a highly significant role in viewing comprehension (Durbahn et al., 2020). This indicates that learners might be motivated to watch L2 television programs if they have gained the adequate receptive vocabulary knowledge to comprehend words in that discourse (Webb & Rodgers, 2009b). Research also shows that L2 television programs are a valuable source of vocabulary learning (e.g., Ahrabi Fakhr et al., 2021; Csomay & Petrović, 2012; Dang, 2020; Fievez et al., 2020; Majuddin et al., 2021; Nguyen & Boers, 2019; Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2011, 2020; Webb & Rodgers, 2009b). Therefore, it is important to have an understanding of the necessary vocabulary knowledge to enjoy L2 television and to explore the likelihood of incidental vocabulary learning through being exposed to this valuable source of linguistic input.

The present study aimed to explore the lexical demands of mainstream scripted and unscripted television programs. Lexical coverage, defined as "the percentage of known words in a text" (Webb, 2010a, p. 498), and frequency of occurrence are two important factors that can shed light on the extent to which scripted and unscripted television programs can be used for language-learning goals. Currently, relatively few studies have addressed the lexical coverage of television programs (Rodgers & Webb, 2011; Webb & Rodgers, 2009b). Furthermore, no study to date has drawn a comparison between scripted and unscripted television programs regarding their vocabulary load. Nation (2016) uses "scripted" and "unscripted" as two different descriptors of spoken discourse. While scripted discourse refers to non-spontaneous, prepared spoken language use, unscripted discourse is characterized by spontaneous, impromptu talk.

Accordingly, scripted television programs are those first written in the form of a script and then verbally articulated by actors, while unscripted ones refer to those in which all talk is spontaneous. However, it should be noted that research has shown scripted television dialogue bears some similarities to that of natural conversation (e.g., Al-Surmi, 2012; Davies, 2021; Quaglio, 2009). The present study is an extension of Webb and Rodgers (2009b), making a comparison between scripted and unscripted television shows with regard to lexical demands. Such a comparison can cast light on the vocabulary level required to achieve adequate comprehension of each type of television show. This study also attempted to address another gap in the literature by conducting a corpus-driven comparison of both types of television programs regarding their potential for incidental vocabulary learning. Our findings may help researchers, teachers, and learners establish vocabulary learning goals that can be achieved through watching both types of television programs.

## Literature Review

### *Vocabulary Knowledge and Comprehension*

Research on L2 comprehension has tried to determine the coverage required to understand spoken and written texts. Although many factors, such as background information (Leeser, 2007) and deriving word meaning from context (Liu & Nation, 1985), affect comprehension, coverage may be the most influential one (Laufer & Sim, 1985), showing how much vocabulary is needed to understand different types of discourse. Studies have suggested different coverage points necessary for gaining adequate comprehension regarding the type of discourse. Laufer (1989) believes that 95% coverage is needed for reasonable comprehension of a text, while Nation (2006) states that 98% is ideal for understanding written texts. Other estimates include 98% for reading for pleasure (Hirsh & Nation, 1992), 98% for adequate unassisted reading comprehension (Hu & Nation, 2000), 98% for reasonable comprehension of academic texts (Schmitt et al., 2011), and 98% for unsimplified written texts (Laufer, 2013).

Two studies have explored the impact of lexical coverage on listening comprehension. Bonk (2000) reported various levels of comprehension at different coverage figures. Based on the results, he suggested that lexical coverage lower than 75% rarely resulted in good comprehension. Rather, he reported that adequate comprehension occurred frequently at 90%+ coverage levels. Van Zeeland and Schmitt (2013b) found that gaining adequate comprehension required 90% and 95% coverage levels. Different coverage figures resulted in different mean scores (i.e., 7.35 at 90%; 7.65 at 95%; 8.22 at 98%; and 9.62 at 100% coverage levels). Their findings also indicated that the majority of both native and non-native participants achieved sufficient comprehension at the 90% threshold; however, the non-native participants' test scores showed great variation at this coverage level. Their results also demonstrated that the amount of variation in their scores at the 95% threshold was far smaller, suggesting that this might be the most appropriate coverage level with regard to listening comprehension.

Durbahn et al. (2020) carried out a study on the amount of coverage required to reach adequate viewing comprehension. Their findings indicated that a rise in the lexical coverage from 87% to 99% resulted in an increase in the participants' comprehension of a documentary from 62% to 92%. Accordingly, they suggested that the amount of lexical coverage needed for viewing comprehension might be less than that for unassisted reading and advanced listening. They also concluded that although relatively good viewing comprehension might be achieved at the 90% coverage level, higher lexical coverage may lead to better comprehension. They reported that "when only 1% of the running words in the video is unknown,

comprehension will be higher” (Durbahn et al., 2020, p. 26). Overall, they concluded that the lexical coverage required for viewing comprehension is similar to the level needed for informal listening comprehension, as reported by van Zeeland and Schmitt (2013b).

However, it is wrong to assume that a particular degree of coverage would guarantee the same level of comprehension for all learners. Therefore, it is reasonable to take other factors into account along with coverage when exploring the comprehension of audiovisual input.

The most notable difference between reading and listening/viewing is the permanent nature of written input in contrast to the temporary nature of auditory/audiovisual input. While reading a text, there is a chance “to refer back to the lexical items to decode the message” (van Zeeland & Schmitt, 2013b, p. 6). This might not be possible in listening/viewing, which makes receiving and decoding the input a challenge (Tegge, 2017). Van Zeeland and Schmitt (2013b) assert that 95% coverage might be a reasonable coverage for comprehending spoken narrative passages, and they also suggest that 98% can be the target when very high comprehension is needed. Research has also shown that audiovisual input may be easier to comprehend, as it is accompanied by visual elements.

Durbahn et al. (2020) pointed out that the lexical coverage level required for the comprehension of audiovisual input might be lower than that for auditory and written input, as the combination of imagery and audio may provide support to guess the meaning of unfamiliar words from context, expand the knowledge of partially known words (e.g., Peters, 2019), and thus enhance viewing comprehension. Based on their findings, they concluded that imagery plays a role in the comprehension of audiovisual input and that it should be treated as a factor that helps differentiate viewing from listening comprehension. In another study, Rodgers (2018) explored the extent to which the imagery and auditory input found in documentary and narrative television programs co-occur and how this might affect vocabulary learning. The results indicated that in both types of programs analyzed, the degree of co-occurrence of imagery with the words in the audio soundtrack was found to support vocabulary learning. Rodgers (2018) contended that the combination of audiovisual and auditory input may foster both comprehension and vocabulary development. Overall, his findings pointed to the advantage of audiovisual input over auditory and written input in building knowledge of unknown vocabulary, which was attributed to the presence of supportive imagery.

### *Word Frequency Lists*

To analyze the lexical demands of different discourse types, researchers have used word-frequency lists. One example is Nation’s (2004) frequency lists, classifying words according to frequency and range data based on an analysis of the British National Corpus (BNC). As these lists were developed from a primarily British corpus, they contained words that were more representative of British English than of American English. Therefore, their use in studies trying to assess the lexical profile of American discourse could be a limitation (e.g., Webb & Rodgers, 2009b). Later, Nation (2012, 2017) conducted an analysis of the British National Corpus and the Corpus of Contemporary American English (COCA), resulting in the development of the BNC/COCA word lists.

Schmitt and Schmitt (2012) suggested that the traditional frequency-based classification of vocabulary be reassessed. Traditionally, high-frequency vocabulary was thought to contain the 2,000 most frequent word families. Considering such factors as “the amount of vocabulary necessary for English usage” and “dictionary defining vocabulary” (p. 484), they contended that high-frequency vocabulary should contain the 3,000 most frequent word families. They also introduced the term “mid-frequency” to account for word families from the fourth 1,000-word level to the ninth 1,000-word level. Nation and Anthony (2013) approved of this reassessment and proposed that high-frequency vocabulary consist of the 3,000 most frequent (i.e., first 1,000 to third 1,000) word families, as this vocabulary level is required to

reach 95% coverage in most texts. Furthermore, mid-frequency vocabulary is suggested to comprise 6,000 (i.e., fourth 1,000 to ninth 1,000) word families. The rationale for considering this boundary is “because 9,000 word families provide 98% coverage of most texts” (Nation & Anthony, 2013, p. 7).

### *The Vocabulary Load of Different Discourse Types*

A number of studies have been conducted to investigate the vocabulary size needed to reach sufficient and ideal coverage in different discourse types.

With respect to 90% coverage considered to provide sufficient viewing and listening comprehension, the necessary vocabulary knowledge has been reported to consist of the 1,000–2,000 most frequent word families. Accordingly, teacher-selected songs (Tegge, 2017), teacher talk (Horst, 2010), and EAL podcasts (Motamedynia & Nasrollahi Shahri, 2022) were lexically least demanding, requiring 1,000 word families, whereas a great number of other discourse types required the 2,000 most frequent word families to reach 90% coverage: television programs (Webb & Rodgers, 2009b), movies (Webb & Rodgers, 2009a), everyday spoken English (Nation, 2006), songs from the pop charts (Tegge, 2017), general-audience podcasts (Motamedynia & Nasrollahi Shahri, 2022; Nurmukhamedov & Sharakhimov, 2021), academic lectures (Dang & Webb, 2014), rap songs (Tegge & Coxhead, 2021), TED Talks (Nurmukhamedov, 2017), and video games (Rodgers & Heidt, 2021).

Given 95% coverage, the necessary vocabulary knowledge is between the 2,000 and 5,000 most frequent word families. The least lexically demanding discourse types were graded readers (Webb & Macalister, 2013), teacher-selected songs (Tegge, 2017), teacher talk (Horst, 2010), and EAL podcasts (Motamedynia & Nasrollahi Shahri, 2022), as they required the 2,000 most frequent word families. In contrast, a number of discourse types needed the 3,000 most frequent word families: television programs (Webb & Rodgers, 2009b), movies (Webb & Rodgers, 2009a), everyday spoken English (Nation, 2006), songs from the pop charts (Tegge, 2017), general-audience podcasts (Motamedynia & Nasrollahi Shahri, 2022; Nurmukhamedov & Sharakhimov, 2021), and high-school textbooks (Nguyen, 2021). Furthermore, the most lexically demanding genres are novels and newspapers (Nation, 2006), academic spoken discourse (Dang & Webb, 2014), TED Talks (Nurmukhamedov, 2017), and rap songs (Tegge & Coxhead, 2021), all requiring the 4,000 most frequent word families, and video games (Rodgers & Heidt, 2021) with 5,000 word families to reach 95% coverage.

However, regarding 98% coverage, the necessary vocabulary knowledge has been suggested to range from the 3,000 to 10,000 most frequent word families. Accordingly, 3,000 word families accounted for 98% coverage of graded readers (Webb & Macalister, 2013), while to reach the same coverage level in teacher-selected songs (Tegge, 2017) and teacher talk (Horst, 2010), knowledge of 4,000 word families was required. More demanding discourse types included general-audience podcasts (Nurmukhamedov & Sharakhimov, 2021) and high-school textbooks (Nguyen, 2021) with 5,000 word families; movies (Webb & Rodgers, 2009a) and songs from the pop charts (Tegge, 2017) with 6,000 word families; and television programs (Webb & Rodgers, 2009b) requiring 7,000 word families. The most demanding genres have been suggested to be academic spoken discourse (Dang & Webb, 2014) and TED Talks (Nurmukhamedov, 2017), each requiring 8,000 word families; novels (Nation, 2006) and rap songs (Tegge & Coxhead, 2021), both needing 8,000–9,000 word families; and video games, requiring 10,000 word families (Rodgers & Heidt, 2021) (see Table 1).

Table 1

## Vocabulary Knowledge Required for 95% and 98% Coverage in Various Discourse Types

<b>Discourse type</b>	<b>90%</b>	<b>95%</b>	<b>98%</b>	<b>Study</b>
General-audience podcasts	2,000+ PN/MW/TC/AC	3,000+ PN/MW/TC/AC	5,000+ PN/MW/TC/AC	Nurmukhamedov & Sharakhimov (2021)
Teacher talk	1,000	2,000	4,000	Horst (2010)
EAL podcasts	1,000+ PN/MW/TC/AC	2,000+ PN/MW/TC/AC	-	Motamedynia & Nasrollahi Shahri (2022)
General-audience podcasts	2,000+ PN/MW/TC/AC	3,000+ PN/MW/TC/AC	-	Motamedynia & Nasrollahi Shahri (2022)
Rap songs	2,000+ PN/MW/TC/AC	4,000+ PN/MW/TC/AC	9,000+ PN/MW/TC/AC	Tegge & Coxhead (2021)
TED Talks	2,000+ PN/MW	4,000+ PN/MW	8,000+ PN/MW	Nurmukhamedov (2017)
High-school textbooks	-	3,000+ PN/MW/TC/AC	5,000+ PN/MW/TC/AC	Nguyen (2021)
Video games	2,000+ PN/MW	5,000+ PN/MW	10,000+ PN/MW	Rodgers & Heidt (2021)
Pop songs	2,000+ PN/MW/TC	3,000+ PN/MW/TC	6,000+ PN/MW/TC	Tegge (2017)
Teacher-selected songs	1,000+ PN/MW/TC	2,000+ PN/MW/TC	4,000+ PN/MW/TC	Tegge (2017)
Spoken academic discourse	2,000+ PN/MW	4,000+ PN/MW	8,000+ PN/MW	Dang & Webb (2014)
Everyday spoken discourse	2,000+ PN	3,000+ PN	7,000+ PN	Nation (2006)
TV programs	2,000+ PN/MW	3,000+ PN/MW	7,000+ PN/MW	Webb & Rodgers (2009b)

Movies	2,000+ PN/MW	3,000+ PN/MW	6,000+ PN/MW	Webb & Rodgers (2009a)
Novels	-	4,000–5,000+ PN	8,000–9,000+ PN	Nation (2006)
Newspapers	-	4,000+ PN	8,000–9,000+ PN	Nation (2006)
Graded readers	-	2,000+ PN/MW	3000+ PN/MW	Webb & Macalister (2013)

*Note.* PN: proper nouns; MW: marginal words; TC: transparent compounds; AC: acronyms

### *Watching Television and Incidental Vocabulary Learning*

Research has explored incidental vocabulary learning through different sources of L2 spoken input (e.g., Çekiç, 2022; Dang et al., 2021; Feng & Webb, 2019; Horst, 2010; Jin & Webb, 2020; Pavia et al., 2019; van Zeeland & Schmitt, 2013a; Vidal, 2003, 2011). Research has also highlighted the potential role that L2 television can play in incidental vocabulary learning (e.g., Ahrabi Fakhr et al., 2021; Csomay & Petrović, 2012; Dang, 2020; Fievez et al., 2020; Majuddin et al., 2021; Nguyen & Boers, 2019; Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2011, 2020; Webb & Rodgers, 2009b). Laufer and Hulstijn (2001) define incidental vocabulary learning as “learning words without deliberate decision to commit information to memory” (p. 11).

One line of research into incidental vocabulary learning has been concerned with the important role that frequency of occurrence plays in vocabulary learning (e.g., Fievez et al., 2020; Majuddin et al., 2021; Peters & Webb, 2018; Rodgers & Webb, 2020; Uchihara et al., 2019; van Zeeland & Schmitt, 2013a; Vidal, 2003, 2011). Frequency of occurrence, though an important predictor of incidental vocabulary acquisition, might not lead to learning (Webb, 2020), as it is one of many factors influencing vocabulary learning. Other factors include background knowledge (Pulido, 2004), prior vocabulary knowledge (Çekiç, 2022; Webb & Chang, 2015; Webb & Paribakht, 2015), L2 proficiency (Zahar et al., 2001), and the amount of information provided in the input (Webb, 2008).

Research findings differ regarding the number of encounters for incidental vocabulary learning to occur. “Learning rarely occurs after a single encounter” (Webb & Macalister, 2013, p. 305), but there is greater potential for learning lexical items when the number of encounters increases (Chen & Truscott, 2010; Webb, 2020). Regarding incidental vocabulary learning through L2 audiovisual sources of input, like television programs and movies, estimates range from at least five encounters for gaining partial knowledge of a word (Rodgers & Webb, 2011; Webb, 2010a; Webb & Rodgers, 2009b) to 10 or more encounters for considering a word a potentially learned item (Rodgers & Webb, 2011; Webb, 2010a). A number of studies have explored the effect of exposure frequency on incidental vocabulary learning through viewing television programs (e.g., Fievez et al., 2020; Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2020), all suggesting that repeated encounters with new words may foster incidental vocabulary learning. Rodgers and Webb (2011) conducted a corpus-based investigation into the potential for the incidental learning of what they considered “low-frequency” vocabulary—that is, Nation’s

(2004) fourth to fourteenth 1,000-word BNC lists—through watching television programs. Their findings revealed that there was greater potential for the incidental learning of such vocabulary through viewing related programs than random ones.

Taken together, research has suggested that to gain an adequate understanding of television programs, “coverage is likely to range from 90% to 99%” (Webb & Rodgers, 2009b, p. 339). This has been supported by Durbahn et al. (2020), who suggested that relatively good viewing comprehension may occur at 90% lexical coverage. They also pointed out that an increase in coverage between 90% and 99% would lead to higher comprehension. Following Durbahn et al. (2020), we chose the 90% coverage level as the threshold to investigate the necessary vocabulary knowledge to gain good comprehension of scripted and unscripted television programs. We also used 95% and 98% coverage figures as other thresholds to explore the lexical knowledge required for higher comprehension of the two types of television shows. These two coverage points have been consistently used in lexical coverage studies to determine the vocabulary size required to gain adequate and ideal comprehension (e.g. Nurmukhamedov & Sharakhimov, 2021; Tegge, 2017; Webb & Rodgers, 2009a, 2009b) and have been supported by a number of studies on the relationship between lexical coverage and reading (e.g., Laufer & Ravenhorst-Kalovski, 2010) and listening comprehension (van Zeeland & Schmitt, 2013b).

The present study also investigated the potential of television programs for incidental vocabulary learning. Webb and Nation (2017) have contended that comprehension of different spoken discourse types requires knowledge of the 3,000 most frequent word families. Moreover, Schmitt and Schmitt (2012) have asserted that mid-frequency words are essential for using English for authentic purposes, comprehension, and proficient language use. Nation (2001) has pointed out that within the limited classroom time, the 1,000 most frequent word families might receive sufficient direct attention, which may not be the case for the 2,000 or 3,000 most frequent word families (Dang & Webb, 2016). Therefore, to reach the lexical goals of gaining knowledge of the 3,000 most frequent word families (necessary to comprehend most spoken discourse types) and the mid-frequency vocabulary (required to comprehend most written discourse types), incidental learning of vocabulary should also be given attention along with intentional learning (Webb, 2020). It might be reasonable, then, to investigate television programs’ potential for the incidental learning of words from the 2,000- to 3,000-word levels, and also mid-frequency vocabulary, to explore whether they are a suitable source of incidental vocabulary learning.

Regarding the necessary frequency of occurrence for incidental vocabulary learning through audiovisual input, we chose the exposure frequency of at least five times as the cut-off, following Rodgers and Webb’s (2020) intervention study on incidental vocabulary learning through viewing television. This frequency of occurrence has also been supported by a number of corpus-based studies on incidental vocabulary learning from watching television shows (Rodgers & Webb, 2011; Webb & Rodgers, 2009b) and movies (Webb, 2010a) as being required to gain partial knowledge of a word. Thus, the present study was guided by the following research questions:

1. What level of vocabulary is required to reach 90%, 95%, and 98% coverage in scripted and unscripted television programs?
2. To what extent do scripted and unscripted television programs hold potential for incidental vocabulary learning?



## Methodology

### *Corpus*

Fourteen different television programs containing 286 episodes and constituting two corpora of 1,698,914 running words were used in this study. The overall viewing time for both corpora was approximately 211 hours. Commercial breaks, if any, were excluded from the scripts and the total running time. The programs were chosen based on three factors: the overall popularity of the programs, genres, and availability of transcripts. The rationale behind the selection process was to choose television programs that were both well-known and popular and that were from common genres. This increased the likelihood of their being watched by a large number of viewers. The programs analyzed in the present study were among the top 150 television shows at the time when the corpora were compiled based on the rankings provided by *IMDB*, the world's most popular and authoritative database of information and statistics about millions of movies, television, and entertainment programs as well as streaming content online, with an average of 550 million monthly visits (Similarweb, 2022). It aims to help enthusiasts explore the world of films and television programs and decide what to watch. It should be noted that the programs were among the most high-ranking shows, with available transcripts, of their own genre at the time of corpus compilation and were classified based on whether they were scripted or unscripted. The scripted corpus included programs from the genres of drama (e.g., *Breaking Bad*) and situation comedy (e.g., *Modern Family*) and contained 849,441 tokens. In contrast, the unscripted corpus included programs from the genres of reality television shows (e.g., *America's Got Talent*) and talk shows (e.g., *Larry King Live*) and consisted of 849,473 words. It is worth mentioning that some of the programs (e.g., *Larry King Live*) were most likely to be watched by adults. It should also be noted that the scripted corpus involved a larger number of programs than the unscripted one, which was due to the wider transcript availability of the former. To ensure that this would not significantly affect the results of the study, we calculated the type/token ratio (i.e., an indicator of lexical diversity) for the two corpora, with the ratios being relatively comparable: scripted corpus = 0.026; unscripted corpus = 0.020. Furthermore, the difference in the number of tokens between the two corpora was kept to a minimum (i.e., 32 words) but that the number of tokens varied among the episodes. More information about the programs can be found in the Appendix.

The transcripts of the programs were downloaded from an online database named *Springfield! Springfield!* (n.d.), which provides scripts of television programs and movies. We manually excluded all of the words that could not be heard in the programs from the transcripts. These included speakers' names, stage directions, scene headings, action sequences, and mood brackets. Following previous lexical profiling studies (e.g., Nurmukhamedov & Sharakhimov, 2021; Webb & Rodgers, 2009a, 2009b), contractions, connected speech, and apostrophized abbreviations were all changed into the standard spelling found in the BNC/COCA word lists. For instance, *don't*, *shoulda*, and *pickin'* were changed into *do not*, *should have*, and *picking*, respectively. If the spellings of such items had not been changed, they would have been classified by the software used for analysis as *not found in any list* words. Webb and Rodgers (2009b, p. 359) state that "knowing the changed spellings does not ensure that you would recognize those items in their original forms". Knowledge of such forms might affect comprehension; however, no research to date has explored the extent to which they can influence comprehension. It is worth noting that although the list of proper nouns in BNC/COCA lists contains a large number of entities (i.e., more than 22,000), there might still be a number of proper nouns in a corpus which this list does not account for. Therefore, those proper nouns that were in the corpora but not in the BNC/COCA proper nouns list were added to this list.

## Analysis

The AntWordProfiler program (Anthony, 2021), version 1.5.1, was used to analyze the corpora. This is a computer program that provides data on the frequency of words and word families in a text. To complete the analysis, the BNC/COCA word lists (Nation, 2012, 2017) were utilized with this software. These frequency lists contain twenty-five 1000-word levels plus four additional lists: proper nouns (e.g., *Alexandra, Robert*), marginal words (e.g., *um, yay*), transparent compounds (e.g., *backdoor, airplane*), and acronyms or initialisms (e.g., *CD, CEO*). It is noteworthy that the word lists contain word families rather than individual words. A word family consists of a headword (e.g., *accident*) and its family members, if any (e.g., *accidents, accidental, accidentally*). Measuring vocabulary knowledge based on word families rests on the assertion that “inflected and regularly derived forms of a known base word can also be considered as known words if the learners are familiar with the affixes” (Hirsh & Nation, 1992, p. 692). Those words that were not found in the 29 word lists were classified by the software as *not found in any list*.

## Results

The first research question was concerned with the vocabulary knowledge required to gain 90%, 95%, and 98% coverage of scripted and unscripted television programs. Table 2 shows the number and percentage of the tokens, that is, the total number of words in a text (Richards & Schmidt, 2002), as well as the number of word types, namely the number of unique word forms (Webb, 2021), and families that occurred at each level in each corpus. Accordingly, the 1,000 most frequent word families accounted for the majority of the tokens, word types, and families in both corpora. The 1,000 most frequent word families accounted for 738,885 (86.98%) and 741,910 (87.34%) of the tokens in the scripted and unscripted programs, respectively. In contrast, the 2,000 most frequent word families accounted for only 31,395 (3.70%) running words in the scripted programs, while the same word level made up 33,565 (3.95%) tokens in the unscripted programs. By the fourth word level, the percentage of the running words is less than 1% in both corpora, highlighting the relative importance of knowing the 3,000 most frequent word families. However, it should be noted that there were still a relatively great number of word families from the fourth 1,000-word level in both corpora: 856 in the scripted and 800 in the unscripted television shows. The difference, however, lies in the number of encounters with the word families at this level, which was far less compared to the first three word levels.

Table 2 also shows that proper nouns accounted for 22,546 (2.65%) and 29,166 (3.43%) of the tokens in the scripted and unscripted programs, respectively. The table also shows that marginal words, transparent compounds, and acronyms accounted for 1.19%/0.76%, 0.35%/0.35%, and 0.10%/0.04% of the tokens in the scripted and unscripted programs, respectively. Nation (2006) proposes that the learning burden of proper nouns and marginal words is not considerable. Moreover, it is plausible to assume that transparent compounds carry a low lexical burden, “as they can also be considered known by means of knowing their high-frequency parts” (Tegge, 2017, p. 89). Nation (2016) also contends that acronyms “should be easy to learn, and their full form should be easy to recall because of the clues given by the initial letters” (p. 85). Following previous lexical coverage studies (e.g., Nurmukhamedov & Sharakhimov, 2021; Tegge, 2017), we added the aggregate percentage of these additional lists to the coverage required to know all the twenty-five 1,000-word lists. It should also be noted that the BNC/COCA lists did not account for 0.29% and 0.13% of the words in the scripted programs and unscripted ones, respectively. These words seem to be from particular sociolects, such as *shawty* and *howdy*; some are technical domain-specific words, such as *cefuroxime* and *lymphocytosis*; some are non-English words, such as *muchas* and *bismillah*; and some tend to be recently coined, such as *funkification* and *exhaustipated*. These words were excluded from the analysis in this study, so the coverage they provided was not taken into account.

Table 2

Tokens, Types, and Word Families at Each Word Level for Both Scripted and Unscripted Programs

word list	scripted				unscripted			
	token (raw)	token (%)	type	word family	token (raw)	token (%)	type	word family
1,000	738885	86.98	3634	999	741910	87.34	3398	997
2,000	31395	3.70	3034	992	33565	3.95	2731	982
3,000	11343	1.34	2200	947	12412	1.46	1968	898
4,000	7828	0.92	1581	856	7494	0.88	1349	800
5,000	5754	0.68	1296	783	3330	0.39	893	614
6,000	3616	0.43	943	625	2852	0.34	631	469
7,000	2764	0.33	731	551	1515	0.18	521	421
8,000	2224	0.26	579	436	1537	0.18	375	306
9,000	1483	0.17	506	408	1285	0.15	318	261
10,000	864	0.10	380	304	549	0.06	236	211
11,000	904	0.11	348	292	594	0.07	202	179
12,000	710	0.08	287	253	788	0.09	167	138
13,000	670	0.08	238	205	451	0.05	103	97
14,000	487	0.06	202	179	292	0.03	87	76
15,000	316	0.04	150	137	184	0.02	73	72
16,000	289	0.03	130	118	153	0.02	57	54
17,000	212	0.02	108	97	83	0.01	40	38
18,000	268	0.03	89	81	117	0.01	42	37
19,000	136	0.02	68	67	108	0.01	26	26
20,000	155	0.02	57	54	97	0.01	33	32
21,000	76	0.01	46	45	18	0.00	13	12
22,000	59	0.01	26	26	23	0.00	12	12
23,000	64	0.01	32	32	28	0.00	12	12
24,000	13	0.00	12	12	20	0.00	5	5

25,000	19	0.00	13	13	6	0.00	4	4
PN	22546	2.65	3519	3455	29166	3.43	2742	2672
MW	10097	1.19	107	34	6487	0.76	88	33
TC	2987	0.35	908	773	2936	0.35	621	525
AC	814	0.10	107	104	337	0.04	86	82
NIL	2463	0.29	1317	1317	1136	0.13	680	680
total	849441		22648	14195	849473		17513	10745

*Note.* PN: proper nouns; MW: marginal words; TC: transparent compounds; AC: acronyms; NIL: not in the lists

Table 3 demonstrates the cumulative coverage for the words with and without proper nouns, marginal words, transparent compounds, and acronyms (i.e. the additional lists) in both corpora. In response to the first research question, Table 3 shows that, without knowledge of the additional lists, the necessary vocabulary sizes to reach 90% and 95% coverage levels were the 2,000 and 11,000 most frequent word families in both corpora, respectively. Furthermore, even vocabulary knowledge of 25,000 word families did not account for 98% coverage in both scripted and unscripted programs. However, if one had knowledge of the additional lists, then a vocabulary size of the 1,000 most frequent word families would be required to achieve 91.92% and 91.24% coverage in the unscripted and scripted programs, respectively. This means that although the two types of programs required the same vocabulary size to reach 90% coverage, the unscripted programs were slightly less lexically demanding than the scripted ones. When the aim was higher coverage levels (i.e., 95% and 98%), however, the required vocabulary sizes varied between the two types of programs. Regarding the 95% threshold, knowledge of 2,000 and 3,000 word families was needed in the unscripted and scripted programs, respectively. Moreover, vocabulary knowledge of the 4,000 most frequent word families plus the additional lists was required in the unscripted programs to achieve 98% coverage, while the figure was 6,000 word families to reach the same coverage point in the scripted ones.

The second research question addressed the potential of the programs, both scripted and unscripted, for incidental vocabulary learning. As the results indicated, knowledge of the 1,000–3,000 most frequent word families was needed to achieve 90%–95% coverage level in both types of television programs. This provides some indication that words that are most likely to be learned incidentally may be those in the second and third 1,000-word levels. In addition, we also explored the programs' potential for the incidental learning of mid-frequency words (i.e., 4,000- to 9,000-word levels), as these have been suggested to provide 98% coverage in most discourse types (Nation & Anthony, 2013; Schmitt & Schmitt, 2012). Considering Rodgers and Webb's (2020) suggestion regarding the required frequency of occurrence for incidental vocabulary learning (i.e., at least five times), the results show that the number of word families from the second and third 1,000-word levels encountered five or more times in the scripted and unscripted television programs was 1,589 and 1,440, respectively. Furthermore, the number of mid-frequency word families encountered five or more times was 1,231 in the scripted programs, whereas the figure for the unscripted programs was 919. Examples of word families from the 2,000- to 3,000- and 4,000- to 9,000-word levels with a frequency of occurrence of five or more in both corpora are *acknowledge* and *congratulate*, respectively.

Table 3

Cumulative Coverage for Both Corpora with and without Proper Nouns, Marginal Words, Transparent Compounds, and Acronyms (i.e., Additional Lists)

word list	scripted		unscripted	
	without additional lists	with additional lists	without additional lists	with additional lists
1,000	86.98	91.24 <sup>a</sup>	87.34	91.92 <sup>a</sup>
2,000	90.68 <sup>a</sup>	94.97	91.29 <sup>a</sup>	95.87 <sup>b</sup>
3,000	92.02	96.31 <sup>b</sup>	92.75	97.33
4,000	92.94	97.23	93.63	98.21 <sup>c</sup>
5,000	93.62	97.91	94.02	98.60
6,000	94.05	98.34 <sup>c</sup>	94.36	98.94
7,000	94.38	98.67	94.54	99.12
8,000	94.64	98.93	94.72	99.30
9,000	94.81	99.10	94.87	99.45
10,000	94.91	99.20	94.93	99.51
11,000	95.02 <sup>b</sup>	99.31	95.00 <sup>b</sup>	99.58
12,000	95.10	99.39	95.09	99.67
13,000	95.18	99.47	95.14	99.72
14,000	95.24	99.53	95.17	99.75
15,000	95.28	99.57	95.19	99.77
16,000	95.31	99.60	95.21	99.79
17,000	95.33	99.62	95.22	99.80

18,000	95.36	99.65	95.23	99.81
19,000	95.38	99.67	95.24	99.82
20,000	95.40	99.69	95.25	99.83
21,000	95.41	99.70	95.25	99.83
22,000	95.42	99.71	95.25	99.83
23,000	95.43	99.72	95.25	99.83
24,000	95.43	99.72	95.25	99.83
25,000	95.43	99.72	95.25	99.83

<sup>a</sup> reaching 90% coverage

<sup>b</sup> reaching 95% coverage

<sup>c</sup> reaching 98% coverage

Table 4

Number of Encounters With 2,000–3,000 and Mid-Frequency Word Families in the Scripted and Unscripted Programs

encounters	scripted		unscripted	
	2,000–3,000-word level	mid-frequency	2,000–3,000-word level	mid-frequency
	word family	word family	word family	word family
1	85	1094	113	1015
2	97	645	110	522
3-4	168	689	217	415
5+	1589	1231	1440	919
total	1939	3659	1880	2871

## Discussion

Regarding the first research question, the results provide some evidence that unscripted television programs might be less demanding than scripted ones regarding lexical burden. The results confirmed this

regarding all of the three thresholds (i.e., 90%, 95%, and 98%) used for analysis in this study. Based on the results, the difference in the lexical demands of the two types of television shows is perfectly clear when the aim is the 95% and 98% coverage levels. Furthermore, although both the scripted and unscripted television programs required the 1,000 most frequent word families to reach 90% coverage, this vocabulary size provided more coverage in the unscripted programs (i.e., 91.92) compared to the scripted ones (i.e., 91.24), suggesting the lower lexical burden of the former. This is corroborated by previous research findings that unscripted discourse may be less demanding than scripted discourse concerning vocabulary load (Adolphs & Schmitt, 2003; Nation, 2006). The present study expanded on the findings of these studies by drawing a comparison between the lexical demands of two types of television programs, which can be unique in their own rights. The overarching finding from this study and earlier ones is that unscripted spontaneous discourse can carry less lexical burden than scripted, previously prepared discourse, and this may also be true in the discourse of television programs as a popular genre. However, it should be noted that the results, though shedding light on the difference between the lexical demands of the discourse in scripted and unscripted television shows, should be treated conservatively, as coverage is one among many factors that influence comprehension. This means that a particular amount of coverage does not equal the same amount of comprehension.

The present study also expanded on previous research into the lexical coverage of television programs (Rodgers & Webb, 2011; Webb & Rodgers, 2009b) as it takes into account unscripted television shows as being different from scripted ones and also explores the lexical demands of television programs at the 90% coverage level. It should be noted that the previous studies used only scripted programs in their analyses and used the 95% and 98% thresholds as their main focus. Regarding 95% coverage, our results regarding the scripted programs are comparable with those of Webb and Rodgers (2009b) and Rodgers and Webb (2011), all suggesting the 3,000 most frequent word families as the required vocabulary size. However, when the target is 98% coverage, there is a discrepancy between the results of the studies. While Webb and Rodgers (2009b) suggest that the 7,000 most frequent word families account for 98% coverage of the programs analyzed in their study, our findings show that to reach the same coverage point the 6,000 most frequent word families are required for the scripted programs. This difference might be due to the different word lists used in the two studies. To analyze their data (both American and British programs), Webb and Rodgers (2009b) used 14 BNC 1,000-word lists developed from a mainly British corpus, which they considered a limitation in their study. However, we utilized 25 BNC/COCA 1,000-word lists, which are derived from both British and American corpora, making them representative of both British and American discourse. Furthermore, the BNC word lists consist of two additional lists: proper nouns and marginal words. In contrast, the BNC/COCA word lists have four additional lists: proper nouns, marginal words, transparent compounds, and acronyms. The added coverage provided by the two extra additional lists (i.e., transparent compounds + acronyms = 0.44%) in this study might be one reason for the difference. Without considering the added coverage provided by these two extra lists, the 6,000 most frequent word families would account for 97.89% coverage, and the 7,000 most frequent word families, as in Webb and Rodgers (2009b), would be needed to reach 98% coverage. Another reason for the difference between the results regarding 98% coverage might be the difference in the makeup of the corpora used in the two studies. The words beyond the third 1,000-word level were shown to add very little at a time to the percentage of known words when the aim is 98% coverage. Therefore, if there is a prominence of a topic, then some of the word lists can make a relatively great contribution to reaching 98% coverage. For example, there was a full season of *The Wire* in the present study, a series with a criminal and legal theme. A separate analysis of the season showed that knowledge of the 2,000 and 5,000 most frequent word families, plus the additional lists, accounted for 95.35% and 98.05% coverage, respectively. This means that the word lists

that contributed to reaching 98% coverage were the third to fifth 1,000-word levels. Results showed that a relatively good number of word families with a criminal and legal theme in the third (e.g., *deputy, jury, jail, prosecute, addict, convict*), the fourth (e.g., *lieutenant, attorney, warrant, squad, juvenile*), and the fifth (e.g. *testify, surveillance, probation, bail, cocaine, heroin*) word levels were found with relatively high frequency in this season. This inflation of such vocabulary resulted in the number of word families required to reach 98% coverage decreasing to the fifth 1,000-word level. This prominence of topic-specific vocabulary in our scripted corpus might be one reason for the discrepancy between our findings and those of Webb and Rodgers (2009b). To potentially avoid this, Webb and Rodgers (2009b) included two episodes per program, which, however, resulted in a much smaller corpus than the one used in the present study.

It is worthwhile to compare the findings not only with previous research on television programs but also with other authentic spoken sources of ESL input. Accordingly, the results showed that both scripted and unscripted television shows may be similar in terms of lexical demands to teacher talk (Horst, 2010), teacher-selected songs (Tegge, 2017), and EAL podcasts (Motamedynia & Nasrollahi Shahri, 2022) and lexically less demanding than other spoken discourse types when the aim is 90% coverage. Furthermore, scripted television programs are similar to some extent to other spoken discourse types regarding 95% coverage. However, teacher-selected songs (Tegge, 2017), EAL podcasts (Motamedynia & Nasrollahi Shahri, 2022), and teacher talk (Horst, 2010) may be less lexically demanding, and rap songs (Tegge & Coxhead, 2021), TED Talks (Nurmukhamedov, 2017), academic spoken discourse (Dang & Webb, 2014), and video games (Rodgers & Heidt, 2021) can be more demanding than scripted television shows regarding lexical load. In contrast, the lexical load of unscripted television programs might be lower than that of most of the other spoken discourse types regarding 95% coverage. Considering the results of this study and those of others, unscripted television programs, teacher-selected songs (Tegge, 2017), and EAL podcasts (Motamedynia & Nasrollahi Shahri, 2022) are perhaps the least lexically demanding spoken discourse types when the aim is 95% coverage, requiring the 2,000 most frequent word families. Regarding 98% coverage, scripted television programs lie somewhere in the middle of the continuum. When the aim is 98% coverage, two discourse types might be less lexically demanding than scripted television programs: general-audience podcasts (Nurmukhamedov & Sharakhimov, 2021) and teacher-selected songs (Tegge, 2017). Furthermore, such television programs are probably lexically similar to pop songs (Tegge, 2017) and movies (Webb & Rodgers, 2009a) with respect to 98% coverage. However, there are some other spoken discourse types, such as TED Talks (Nurmukhamedov, 2017), rap songs (Tegge & Coxhead, 2021), and academic lectures (Dang & Webb, 2014) that might be more demanding than scripted television shows. Furthermore, the vocabulary size necessary to reach 98% coverage in unscripted television shows is the same as that in teacher-selected songs (Tegge, 2017) but smaller than other spoken discourse types. This, therefore, might show that television programs might be a valuable source of mid-level authentic materials to improve listening comprehension and vocabulary learning.

Although there is a need for increased viewing time and therefore more encounters to consider a word potentially learned, a smaller number of encounters can still provide the conditions for a partial grasp of word knowledge (Webb, 2007). Research on television's potential for incidental vocabulary learning suggests that encountering a word at least five times might result in gaining partial word knowledge (Rodgers & Webb, 2011, 2020; Webb, 2010a; Webb & Rodgers, 2009b). Given the results of this study, both types of programs might hold relatively great potential for learning words from the 2,000- to 3,000-word levels, as 1,589 and 1,440 of the word families from these word levels were encountered at least five times in the scripted and unscripted programs, respectively. Furthermore, our findings indicated that both the scripted (1,231 word families) and unscripted (919 word families) programs might hold some potential for the incidental learning of mid-frequency vocabulary. The results also indicated that scripted programs tend to account for a greater number of word families encountered five or more times from both the 2,000- to 3,000- and 4,000- to 9,000-word levels compared to the unscripted ones (i.e., 2,000–3,000: 1,589 vs. 1,440; 4,000–9,000: 1,231 vs. 919). This provides some indication that scripted programs might have a greater



potential for the incidental learning of words from the second to third and fourth to ninth 1,000-word levels than unscripted ones. This suggests that scripted programs might also be more suitable than unscripted ones when the aim is the incidental learning of such words. Taken together, the results attest to L2 television's potential, in both scripted and unscripted programs, for the incidental learning of words from the 2,000- to the 9,000-word levels.

It should also be noted that some of the word families were found across the corpora and were not limited to one program. For example, the word family *congratulate* occurred 97 times (range: 62) across the scripted corpus and 327 times (range: 71) across the unscripted corpus and could be found in all of the programs analyzed. However, there were a number of word families that were not equally distributed across each of the corpora, most of which were semantically domain-specific. For instance, the word family *tumour* occurred 87 times in 17 different episodes across the scripted corpus, with 80 occurrences being in 12 episodes of *House MD*, which is a series with a medical theme. Another example is *salmon*, with a frequency of 123 in the unscripted corpus, 119 of which were in different episodes of *Hell's Kitchen*, a cooking-themed reality television show. This might suggest that watching several episodes of the same program may provide a chance for learning domain-specific vocabulary, while words that are not domain-specific might be learned through watching both related and unrelated programs provided that the exposure time is long enough.

Taken together, the results are supported by previous research highlighting television's great potential for incidental vocabulary learning (e.g., Ahrabi Fakhr et al., 2021; Csomay & Petrović, 2012; Dang, 2020; Fievez et al., 2020; Majuddin et al., 2021; Nguyen & Boers, 2019; Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2011, 2020; Webb & Rodgers, 2009b). Although television's potential for vocabulary learning was considered in terms of frequency of occurrence in the present study, it should be noted that other factors may affect this process, including background knowledge (Pulido, 2004), learners' prior vocabulary knowledge (Çekiç, 2022; Webb & Chang, 2015; Webb & Paribakht, 2015), L2 proficiency (Zahar et al., 2001), and the amount of information provided in the input (Webb, 2008). This means that encountering a word a certain number of times does not guarantee that it will be learned incidentally, so the results should be interpreted with caution.

## **Pedagogical Implications**

Research has attested to television's value as an authentic source of L2 input (e.g., Lindgren & Muñoz, 2013; Peters, 2018; Rodgers & Webb, 2020; Webb, 2015) and of vocabulary learning (e.g., Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2011, 2020; Webb & Rodgers, 2009b). Regarding our findings on the necessary vocabulary levels for achieving good and adequate comprehension of television programs, it should be noted that gaining knowledge of the 1,000 and 2,000–3,000 most frequent word families, though a difficult task for many EFL learners (Webb & Nation, 2017), might be an achievable goal for most ESL learners, as suggested by Webb and Rodgers (2009a, 2009b). However, it should be borne in mind that learning the most frequent 1,000 may be an appropriate initial vocabulary learning target, as our findings showed that this amount of vocabulary knowledge accounted for 90% coverage of scripted and unscripted television programs, which might be enough to achieve relatively good comprehension. Once learners have gained knowledge of the 1,000 most frequent word families, it might be reasonable to establish the more ambitious goal of learning the 2,000–3,000 most frequent word families, because such vocabulary sizes were found to account for the higher coverage of 95% level in the television programs analyzed in this study. However, acquiring knowledge of the 4,000–6,000 most frequent word families

suggested to be required to reach 98% coverage in television programs might be an unrealistic goal, if not unattainable, for a great many learners. Although reaching a coverage level as high as 98% may be ideal (Webb & Rodgers, 2009b), lower coverage points of 95% (Rodgers & Webb, 2011; Webb & Rodgers, 2009b) and even 90% (Durbahn et al., 2020) may also be adequate to promote comprehension as well as vocabulary learning. Both teachers and learners should bear in mind that higher coverage is likely to facilitate comprehension and incidental vocabulary learning (Webb & Rodgers, 2009b). However, learners' proficiency levels should be given careful attention, as it is an important factor. If learners with very low proficiency levels watch L2 television, their comprehension might be too low, which may result in their being discouraged from further viewing (Webb, 2015). Based on the findings of the present study and from a learning opportunity perspective, learners with lower levels of vocabulary knowledge might stand to benefit more from unscripted television programs, which were found to impose a smaller lexical burden compared to scripted ones. However, it should be noted that learners' interests and preferences might be a more significant factor when selecting a television program because it might lead to more viewing when compared with programs that provide learning opportunities alone. This is important because learners ought to be interested in the program to be motivated to watch it (Rodgers, 2016). If learners watch the programs they find interesting, then it is more likely that they receive the large amounts of audiovisual input required for language learning gains (Rodgers, 2018). Furthermore, higher proficiency learners with greater vocabulary knowledge can be encouraged to watch both unscripted and scripted programs because their comprehension may be at the appropriate level.

The fact that watching television in a second or foreign language can be a pleasurable leisure activity could make it a useful tool for enhancing vocabulary knowledge. Research suggests large incidental L2 vocabulary gains through watching television (e.g., Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2013; Rodgers & Webb, 2011, 2020). An increase in lexical gains can occur through learning new vocabulary, strengthening learners' knowledge of partially known words, and acquiring knowledge of the spoken forms of words (Webb, 2015). Webb and Rodgers (2009b) assert that with knowledge of the 3,000 most frequent word families, large gains in vocabulary knowledge through incidental learning can be expected if learners regularly watch L2 television over a long period of time. Therefore, attempts should be made to encourage learners to gain adequate knowledge of the 3,000 most frequent word families in order to pave the way for incidental vocabulary learning. Vilkaitė-Lozdienė and Schmitt (2020) contend that "high-frequency words are the most useful ones and they give learners the best value for their study effort. Thus, they need special attention in a language classroom" (p. 82). Consequently, it seems plausible to assert that high-frequency words should be given special attention in an extensive viewing program through television programs, as they often provide a "95% coverage level for many texts" (p. 87). Furthermore, Webb and Nation (2017) suggest that mid-frequency words (i.e., the 4,000- to 9,000-word levels) are important because they are encountered frequently enough to support comprehension. The results of the present study showed that both scripted and unscripted television programs might have relatively good potential for the incidental learning of words from the 2,000- to 3,000- (i.e., two levels of high-frequency vocabulary) as well as the 4,000- to 9,000-word levels (i.e., mid-frequency vocabulary). Therefore, when the target is the incidental learning of words from the second to the ninth 1,000-word levels, learners might benefit from watching both types of programs, with scripted ones creating more incidental learning opportunities. It should also be noted that the amount of viewing time associated with the potential for the vocabulary gains found in this study is very large (i.e., nearly 211 hours), which may not be possible in a classroom-based program. However, if learners are motivated to autonomously watch L2 television programs on a regular basis over a long period of time, which Webb (2015) considers a principle of an extensive viewing program, then the vocabulary gains suggested in the present study may occur.

Research has suggested that an increase in the number of encounters with a word would increase its chance of being learned (e.g., Peters & Webb, 2018; Webb, 2007). This might provide some evidence that extensive viewing of television programs could increase the number of encounters with words and

therefore facilitate vocabulary learning. According to Webb (2015), extensive viewing is characterized by uninterrupted in-class and out-of-class watching of L2 television on a regular basis. Teachers can make use of television programs' potential, both scripted and unscripted, for incidental vocabulary learning to establish extensive viewing programs for their learners, in which they "would be encouraged to choose television programs that interest them and to view successive episodes of these programs" (Rodgers & Webb, 2020, p. 213). Thanks to advances in information technology and the widespread availability of internet streaming services, EFL learners could be encouraged to participate in extensive viewing programs to increase their viewing time and thus increase their vocabulary gains via watching scripted and unscripted television shows.

It is also noteworthy that watching single episodes of television programs might not provide learners with enough encounters with mid-frequency vocabulary, making the incidental learning of such words quite difficult. However, watching single episodes might facilitate vocabulary learning provided that learners watch them over and over again (Webb & Rodgers, 2009b). Research also suggests that learners might benefit more from watching related television programs compared to watching unrelated television programs with respect to better comprehension and larger vocabulary gains (Rodgers & Webb, 2011). According to the results of the present study, learners might have a greater chance of learning domain-specific words if they watch different episodes of the same programs. This means that learners should be encouraged to watch different episodes of the same program, either scripted or unscripted, as it may increase the chance of certain word families reoccurring. Comprehension and vocabulary learning are also suggested to increase if learners watch programs with similar topics and storylines (Webb, 2010b; Webb & Rodgers, 2009b).

## Conclusion

This study provides insight into the lexical demands of scripted and unscripted television programs and their potential for incidental vocabulary learning. Our findings have implications regarding the importance of investigating the lexical loads of these two types of television programs for language learning and teaching to ensure their appropriateness for the intended learners and intended use. This study also provides some evidence that both scripted and unscripted television programs might have relatively good potential for the incidental learning of words from the 2,000- to the 9,000-word levels.

There were, however, a number of limitations in the present study which should be taken into account. Researchers, teachers, and learners should bear in mind that vocabulary knowledge is one of many factors influencing comprehension. Factors such as background knowledge (Leeser, 2007) and L2 proficiency (Zahar et al., 2001) may also influence both comprehension and vocabulary learning. Moreover, the idea of using word families as a meaningful unit for calculations of receptive knowledge has been challenged. For example, McLean (2017) has proposed that knowledge of a base word does not necessarily result in knowing all of its word family members. Furthermore, a number of the programs analyzed in this study are unlikely to be watched by children and even teenagers. Therefore, the findings should not be generalized, as the television shows popular among these groups of learners might have different lexical demands. Furthermore, the corpus used in this study involved a smaller number of unscripted programs than scripted ones. Even though the type/token ratio for each of the two sub-corpora was relatively similar, the unequal selection of scripted and unscripted shows must be taken into consideration when interpreting the findings of the study. In addition, the television programs were from a variety of genres, suggesting that each represents a different semantic domain affecting the vocabulary covered, meaning that the

findings of the present study should be interpreted with caution. Although the findings provide insights into the lexical demands of scripted and unscripted television programs, teachers should assess the vocabulary level of programs, and even of single episodes, separately before assigning them to their learners. Finally, despite the relatively large corpus used in this study, future research can further investigate the lexical coverage of the two types of programs by analyzing far larger corpora.

### *The Authors*

Hesamoddin Shahriari is a professor at Centennial College, Canada. His research is focused on exploring formulaic language and its implications for language teaching in general and academic English in particular.

Masoud Motamedynia has a master's degree in applied linguistics from the University of Tehran, Iran. His research interests include applied corpus linguistics, vocabulary learning and teaching, discourse analysis, and register analysis.

### *References*

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425–438. <https://doi.org/10.1093/applin/24.4.425>
- Ahrabi Fakhr, M., Borzabadi Farahani, D., & Khomeijani Farahani, A. A. (2021). Incidental vocabulary learning and retention from audiovisual input and factors affecting them. *English Teaching & Learning*, 45, 167–188. <https://doi.org/10.1007/s42321-020-00066-y>
- Al-Surmi, M. (2012). Authenticity and TV shows: A multidimensional analysis perspective. *TESOL Quarterly*, 46(4), 671–694. <https://doi.org/10.1002/tesq.33>
- Anthony, L. (2021). *AntWordProfiler* (Version 1.5.1) [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software>
- Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14(1), 14–31. <https://doi.org/10.1080/10904018.2000.10499033>
- Çekiç, A. (2022). Incidental L2 vocabulary learning from audiovisual input: The effects of different types of glosses. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2022.2062004>
- Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31(5), 693–713. <https://doi.org/10.1093/applin/amq031>
- Csomay, E., & Petrović, M. (2012). “Yes, your honor!”: A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System*, 40(2), 305–315. <http://dx.doi.org/10.1016/j.system.2012.05.004>
- Dang, T. N. Y. (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *TESOL Quarterly*, 54(2), 436–459. <https://doi.org/10.1002/tesq.552>
- Dang, T. N. Y., Lu, C., & Webb, S. (2021). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*, 44(3), 708–736. <https://doi.org/10.1017/S0272263121000474>

- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76. <https://doi.org/10.1016/j.esp.2013.08.001>
- Dang, T. N. Y., & 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>
- Davies, M. (2021). The TV and Movies corpora: Design, construction, and use. *International Journal of Corpus Linguistics*, 26(1), 10–37. <https://doi.org/10.1075/ijcl.00035.dav>
- 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>
- Feng, Y., & Webb, S. (2019). Learning vocabulary through reading, listening, and viewing. *Studies in Second Language Acquisition*, 42(3), 499–523. <https://doi.org/10.1017/S0272263119000494>
- Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2020). Vocabulary learning through viewing captioned or subtitled videos and the role of learner- and word-related factors. *Computer Assisted Language Learning*, 37(3), 233–253. <https://doi.org/10.1558/cj.39370>
- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689–696.
- Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language*, 22(1), 161–180. <https://doi.org/10.125/66646>
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal*, 104(3), 550–566. <https://doi.org/10.1111/modl.12661>
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.
- Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *TESOL Quarterly*, 47(4), 867–872. <https://doi.org/10.1002/tesq.140>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Laufer, B., & Sim, D. D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals*, 18(5), 405–411. <https://doi.org/10.1111/j.1944-9720.1985.tb00973.x>
- Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, 57(2), 229–270. <https://doi.org/10.1111/j.1467-9922.2007.00408.x>
- Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 10(1), 105–129. <https://doi.org/10.1080/14790718.2012.679275>
- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16(1), 33–42. <https://doi.org/10.1177/003368828501600103>

- Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, 43(5), 985–1008. <https://doi.org/10.1017/S0272263121000036>
- McLean, S. (2017). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39(6), 823–845. <https://doi.org/10.1093/applin/amw050>
- Motamedynia, M., & Nasrollahi Shahri, 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. <https://doi.org/10.37213/cjal.2022.32746>
- Nation, I. S. P. (2001). How many high frequency words are there in English? In M. Gill, A. W. Johnson, L. M. Koski, R. D. Sell, & B. Wärvik (Eds.), *Language, learning, literature: Studies presented to Håkan Ringbom* (English Department Publications, vol. 4, pp. 167–181). Åbo Akademi University. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/2001-High-freq-Hakan.pdf>
- Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). John Benjamins. <https://doi.org/10.1075/llt.10.03nat>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2012). The BNC/COCA word family lists. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.
- Nation, I. S. P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>
- Nation, I. S. P., & Anthony, L. (2013). Mid-frequency readers. *Journal of Extensive Reading*, 1(1), 5–16. [https://www.lexutor.ca/conc/eng/nation\\_anthony\\_2013.pdf](https://www.lexutor.ca/conc/eng/nation_anthony_2013.pdf)
- Nguyen, C. D. (2021). Lexical features of reading passages in English-language textbooks for Vietnamese high-school students: Do they foster both content and vocabulary gain? *RELC Journal*, 52(3), 509–522. <https://doi.org/10.1177/0033688219895045>
- Nguyen, C. D., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED Talk. *TESOL Quarterly*, 53(1), 5–29. <https://doi.org/10.1002/tesq.441>
- Nurmukhamedov, U. (2017). Lexical coverage of TED Talks: Implications for vocabulary instruction. *TESOL Journal*, 8(4), 768–790. <https://doi.org/10.1002/tesj.323>
- Nurmukhamedov, U., & Sharakhimov, S. (2021). Corpus-based vocabulary analysis of English podcasts. *RELC Journal*. Advance online publication. <https://doi.org/10.1177/0033688220979315>
- Organisation for Economic Co-operation and Development (OECD). (2009). *Communications Outlook 2009*. OECD.
- Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, 41(4), 745–768. <https://doi.org/10.1017/S0272263119000020>
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *International Journal of Applied Linguistics*, 169(1), 142–168. <https://doi.org/10.1075/itl.00010.pet>
- 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., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148. <https://doi.org/10.1016/j.system.2016.10.002>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577. <https://doi.org/10.1017/S0272263117000407>
- Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, 54(3), 469–523. <https://doi.org/10.1111/j.1467-9922.2007.00415.x>
- Quaglio, P. (2009). *Television dialogue: The sitcom Friends vs. natural conversation*. John Benjamins.
- Richards, J. C., & Schmidt, R. (2002). Token. In *Longman Dictionary of Language Teaching & Applied Linguistics* (3<sup>rd</sup> ed., p. 555). Longman.
- Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* [Doctoral dissertation, Victoria University of Wellington, Wellington]. <https://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/2870/thesis.pdf?sequence=2>
- Rodgers, M. P. H. (2016). Extensive listening and viewing: the benefits of audiobooks and television. *The European Journal of Applied Linguistics and TEFL*, 5(2), 43–57.
- Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words. *International Journal of Applied Linguistics*, 169(1), 192–213. <https://doi.org/10.1075/itl.00012.rod>
- Rodgers, M. P. H., & Heidt, J. (2021). Levelling up comprehensible input and vocabulary learning: The lexical profile of video games. In V. Werner & F. Tegge (Eds.), *Pop culture in language education: Theory, research, practice* (pp. 215–227). Routledge.
- Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, 45(4), 689–717. <https://doi.org/10.5054/tq.2011.268062>
- Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *International Journal of Applied Linguistics*, 171(2), 191–220. <https://doi.org/10.1075/itl.18034.rod>
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Similarweb. (2022, July). *imdb.com*. <https://www.similarweb.com/website/imdb.com/#ranking>
- Springfield! Springfield! (n.d.). *TV show episode scripts*. <https://www.springfieldspringfield.co.uk/>
- Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System*, 67, 87–98. <https://doi.org/10.1016/j.system.2017.04.016>
- Tegge, F., & Coxhead, A. (2021). Exploring the vocabulary of rap lyrics. In V. Werner & F. Tegge (Eds.), *Pop culture in language education: Theory, research, practice* (pp. 71–84). Routledge.
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>

- van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41(3), 609–624. <https://doi.org/10.1016/j.system.2013.07.012>
- van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage and 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>
- Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24(1), 56–89. <https://doi.org/10.1093/applin/24.1.56>
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258. <https://doi.org/10.1111/j.1467-9922.2010.00593.x>
- Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 81–95). Routledge.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65. <https://doi.org/10.1093/applin/aml048>
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245. <http://www2.hawaii.edu/~readfl/rfl/October2008/webb/webb.pdf>
- Webb, S. (2010a). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, 15(4), 497–519. <https://doi.org/10.1075/ijcl.15.4.03web>
- Webb, S. (2010b). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, 11(1), 117–135. <https://doi.org/10.6018/ijes/2011/1/137131>
- Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge.
- Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Eds.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.
- Webb, S. (2021). The lemma dilemma: How should words be operationalized in research and pedagogy? *Studies in Second Language Acquisition*, 43(5), 941–949. <https://doi.org/10.1017/S0272263121000784>
- Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, 37(4), 651–675. <https://doi.org/10.1017/S0272263114000606>
- Webb, S., & Macalister, J. (2013). Is text written for children useful for L2 extensive reading? *TESOL Quarterly*, 47(2), 300–322. <https://doi.org/10.1002/tesq.70>
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- 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, 34–43. <https://doi.org/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. <https://doi.org/10.1093/applin/amp010>
- Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366. <https://doi.org/10.1111/j.1467-9922.2009.00509.x>
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, 57(4), 541–572. <https://doi.org/10.3138/cmlr.57.4.541>



This work is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



The TESL Canada Journal retains copyright of this work. The TESL Canada Journal, as the original source of publication, along with the original author(s), must be acknowledged in any reuse or adaptation of published material. Reuse includes distribution, adaptation, and building upon the material in any medium or format. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.

### *Appendix: Lists of Programs*

#### **Scripted programs**

1. *Breaking Bad* (20 episodes)
2. *Game of Thrones* (20 episodes)
3. *House, M.D.* (22 episodes)
4. *The Wire* (13 episodes)
5. *30 Rock* (21 episodes)
6. *Curb your Enthusiasm* (10 episodes)
7. *Glee* (22 episodes)
8. *Modern Family* (24 episodes)
9. *Suits* (16 episodes)
10. *Veep* (10 episodes)

#### **Unscripted programs**

11. *America's Got Talent* (24 episodes)
12. *Hell's Kitchen* (15 episodes)
13. *The Voice (US)* (29 episodes)
14. *Larry King Live* (40 episodes)