


Binge-watching TV series: English subtitles as an alternative to graded readers for extensive reading

Wenhua Hsu 

I-Shou University, No. 1, Sec. 1, Syuecheng Rd, Dashu District, Kaohsiung City 84001, Taiwan
whh@isu.edu.tw



ARTICLE INFO

Article history

Received 16 August 2023
Revised 12 October 2023
Accepted 15 December 2023

Keywords

English subtitles
Extensive reading
Vocabulary level

ABSTRACT

'Drama fever' has been riding high with increasing consumer usage of OTT streaming services and prevalence of Internet-connected mobile devices, leading to the phenomenon of binge-watching on college campuses in Taiwan. In lieu of graded readers, which top out at the first 3000–4000 word-family levels, this study targeted English subtitles as an alternative source of input, since they also offer EFL learners a channel for exposure to English. To examine this alternative, this study first measured the vocabulary levels of English subtitles for the potential thereof as extensive reading material. The researcher compiled four corpora with each having approximately 2.5 million English-subtitled words from Korean, Japanese, American and Chinese TV series across a couple of genres with high viewership ratings on OTT services for comparison. The operational measures involved the ranked twenty-five 1000-word-family lists along the word-frequency scale of the British National Corpus and the Corpus of Contemporary American English. Results showed that English-subtitled Korean, Japanese and Chinese dramas reached the 2nd–3rd 1000 word-family levels at 95% text coverage and the 4th–5th 1000 levels at 98% coverage, while American series extended to the 7th–8th 1000 levels at 98% coverage from the 3rd–4th 1000 levels at 95% coverage. Moreover, ten participants expressed their views on watching English-subtitled dramas to a certain level of agreement. The data may serve as a reference concerning the vocabulary goal within the first 5000 word families for EFL learners if they continually binge-watch drama series in their leisure time.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



How to Cite: Hsu, W. (2023). Binge-watching TV series: English subtitles as an alternative to graded readers for extensive reading. *English Language Teaching Educational Journal*, 6 (3), 151-161. <https://doi.org/10.12928/eltej.v6i3.10068>

1. Introduction

In recent years, how people watch television (TV) series has changed a lot. With the rapid rise of over-the-top (OTT) streaming services (such as Netflix, Disney+, Hulu, Amazon Prime Video), the shift has been significant from traditional appointment viewing to viewing on streaming services, especially for young people. Appointment viewing means sitting on the couch in front of a television set to watch satellite or cable TV programs in a once-daily or once-weekly broadcasting format, while streaming platforms via Internet-connected mobile devices can be accessed anywhere anytime, which allows for binge-watching with much ease. According to Cambridge Dictionary, the meaning of binge-watch is to watch several episodes of a TV series or program one after another. In a survey carried out by Netflix (2014) on viewing habits, approximately 1500 TV streamers described their binge-watching behavior as viewing from two to six episodes in one sitting.

This research motive stems from the binge-watching phenomenon on college campuses in Taiwan, where English is a foreign language (EFL) and a required subject. During the class break, the researcher at times overhears her students chatting about which drama is streaming and popular. Binge-watching dramas seems to have become one of their leisure activities in free time. On OTT platforms such as Netflix, where drama series are broadcast to global audiences, English is often the first language into which foreign dramas are translated, and then as a pivot translation, English subtitles serve as the source texts for translation into many other languages (Locher & Messerli, 2020; Pedersen, 2019). However, a regularly repeated, informal classroom survey over the past year told the researcher that an overwhelming majority of our students primarily chose Mandarin (the first language) subtitles to watch. The few chose English subtitles are English majors. Right after the question concerning which language subtitles are used is the quick inquiry about which language drama they prefer to watch. Nearly three-fifths of the students being asked to answer admitted that they are Korean drama buffs, while one-fifth of them expressed their preference for Japanese dramas and less than one-fifth for Chinese dramas. Only a handful of students stated that they like to watch American movies rather than American TV series on OTT platforms, but they use Mandarin subtitles. As such, the researcher has been concerned with English lexical growth if the viewing habit shifts from Mandarin to English subtitles.

Subtitles are on-screen texts in the viewer's first language (L1) as opposed to L2 captions in L2 videos (Markham & Peter, 2003). Different from the two standard situations (L1 subtitles and L2 captions in L2 videos), the current context refers to subtitles in the viewer's L2 (English) in L3 (Korean, Japanese) videos, which may compel Taiwanese students with little knowledge of L3 to rely upon on-screen texts in English if they do not use Mandarin subtitles. In this situation, English subtitles in L3 dramas may be compared to graded readers that English teachers like ours often encourage their EFL students to read in an extensive reading scheme.

Vocabulary learning is a gradual process. Extensive reading, informed by Krashen's (1985) input hypothesis, has long been promoted for vocabulary learning (Day & Bamford, 1998, 2004; Horst, 2005). Learners are encouraged to choose reading materials of their interest that are lower than or equivalent to their proficiency level without feeling daunted (Waring & McLean, 2015). As the word 'extensive' has suggested, learners need to engage in voluminous reading in order to have sufficient encounter with unfamiliar words for acquisition to occur. They need to read in quantities large enough to eventually internalize how those unfamiliar words are used and how they collocate with the words surrounding them (Nation, 2015).

To carry out an extensive reading scheme, English teachers often adopt reputable publishers' vocabulary-controlled graded reader series (e.g. Cambridge Readers, Oxford Bookworms, Macmillan Guided Readers, Penguin Readers) for their students to choose from. Having six to seven levels and ranging from 250 to 3800 headwords, graded readers top out at the first 3000–4000 word-family levels (Claridge, 2012; McQuillan, 2016). As the vocabulary levels have shown, graded readers can be an important source of input for intermediate-level EFL learners. However, due to a limited budget for purchase, the number of such readers that EFL learners are asked to read may be actually still small. There is also the question of interest—not every EFL learner thinks of graded readers to be engaging enough to do voluminous reading, which entails reading several million words.

To enhance English vocabulary growth, Webb and Rodgers (2009) advocated that watching English TV programs may be intriguing to L2 learners. They collected eighty-eight English scripts of various American and British TV series and gauged the vocabulary levels thereof along the word-frequency scale of the British National Corpus, which ranks words from the 1st to 14th 1000-word-family levels. Using 95% and 98% lexical coverage of text (knowledge of 95 and 98 words per 100 words) as the lower and upper thresholds, they measured the vocabulary demands of TV programs. Results showed that knowledge of the first 2000 to 4000 word families plus proper nouns and marginal words would account for 95% coverage of TV program transcripts, whereas knowledge of the first 5000 to 9000 word families plus proper nouns and marginal words would provide 98% coverage subject to drama genres.

In subsequent research on how to effectively use TV programs for English learning, Rodgers and Webb (2011) recommended narrow viewing by watching programs with related themes rather than random episodes from programs with different themes or by watching a series of seasons of programs with similar storylines. They offered some evidence that the number of words occurring once was

larger and the number of low-frequency words occurring ten times or more was smaller in unrelated TV programs than in related programs. This implies that TV series with a story arc linking episodes thereof are likely to use fewer variety of words and thus have a lower vocabulary load than unrelated programs, yet for learners, the chance of having a greater number of encounters with those words is higher, increasing the likelihood of vocabulary learning.

Using American TV shows streaming on Netflix, Dizon (2016) compared intermediate-level Japanese EFL learners' perceptions towards the use of L1 subtitles, L2 subtitles, and no subtitles. Among the three subtitling modes, L2 subtitles were favored over L1 subtitles or no subtitles, suggesting that the learners neither would like to depend on their mother tongue nor had the English proficiency enough to watch the video without subtitles. In a subsequent study, Dizon (2018) provided nine Japanese EFL students with free access to Netflix for three months in a bid to find out how these students use the on-demand video resources in their spare time. Most comments collected from these students mentioned the benefits of L2 subtitles for comprehension and vocabulary learning.

Through a scoping review and meta-analysis, Reynolds, Cui, Kao and Thomas (2022) scrutinized 139 past studies in an attempt to understand how the viewing of captioned and subtitled videos can lead to effective vocabulary learning. They concluded that viewing captioned or subtitled videos in any type had a positive effect on vocabulary learning. Reynolds et al's (2022) scoping review has some implications for this research in terms of vocabulary learning in L3 video or L1 video with L2 (target language) subtitles.

As mentioned above, avid drama fans like our students can marathon a TV series in one sitting during holidays or engage in "serial viewing" (Rubenking & Bracken, 2021, p. 100356) over the course of several days or weeks in their spare time. If Taiwanese college students use English (L2) subtitles in lieu of Mandarin (L1) subtitles, the researcher is inclined to believe that the impetus may make possible sustained reading of English subtitles while binge-watching TV series on OTT. As with graded readers, English subtitles in this scenario turn to be lexical learning resources of value. If we draw parallels between extensive reading of graded readers and voluminous reading of English subtitles, they both share some common purposes such as reading for pleasure, building reading fluency as well as sustaining vocabulary growth (Day & Bamford, 2004).

This research addressed two questions concerning English subtitles as an alternative to graded readers for extensive reading.

1. What are the vocabulary levels of English-subtitled Korean, Japanese and Chinese dramas as opposed to American dramas?
2. What are the learners' perceptions regarding viewing English-subtitled dramas?

2. Methodology

2.1. The Corpora

The present corpora consisted of four sub-corpora (Korean, Japanese, American and Chinese dramas), each containing approximately 2.5 million words from different TV series across a couple of genres (e.g., romance, coming of age, action, crimes, fantasy, thriller, legal and medical dramas) with high viewership ratings on OTT services, totaling 10,009,068 running words (see Table 1). Inclusion of popular dramas in each sub-corpus was based upon the assumption that the TV series of high viewership on OTT platforms are likely to have been watched by a large number of drama fans around the world. It is worth highlighting that most TV series are a fusion of several genres rather than limited to a single genre. For example, a medical series may be intertwined with genres such as crimes, romance or even comedy, leading to many twists and turns, say, homicidal scenarios and legal interrogations. In terms of corpus size (over 10 million words), genre balance and high viewership, the present corpus should suffice to provide a reliable measurement of vocabulary levels thereof.

TV series English subtitles in SubRip Text (SRT) files were downloaded from the Internet for research purposes. They are the plain text files, including the subtitles in sequence along with the start and end timecodes. There are one or two lines of subtitles per scene with 7 to 8 words at most per line. Below is the excerpt in SRT format from Episode 1 of *Crash Landing on You*, a blockbuster series on Netflix Asian in 2019—2020, for a snapshot of SRT files.

70
00:05:50,200 --> 00:05:52,240
-Damn it.
-Did she hang up on you again?
71
00:05:52,330 --> 00:05:54,660
-This little brat.
-Just leave her be.
72
00:05:54,750 --> 00:05:57,580
It's been years since Se-ri estranged
the whole family. Don't bother.
73
00:05:57,670 --> 00:05:59,960
I only called her
because Dad wanted me to.
74
00:06:00,880 --> 00:06:03,630
By the way, why did Father
summon all of us today?
75
00:06:04,380 --> 00:06:05,720
That must be it, right?
76
00:06:06,800 --> 00:06:09,800
-Announcing the heir.
-Announcing the heir.

Table 1. The English Subtitle Corpus

TV series with high viewership ratings		Number of series	Total words
Korean (L3) dramas	Squid Game, The Glory (2 seasons), the World of the Married, Reborn Rich, Sky Castle, Crash Landing on You, Reply 1988, Goblin, Doctor Cha, Extraordinary Attorney Woo, Mr. Queen, Crash Course in Romance, Hospital Playlist (2 seasons), Moving, My Dearest (2 seasons), Alchemy of Souls (2 seasons), Dr. Romantic (3 Seasons), Taxi Driver (2 seasons)	25 (Each series ranges from 16 to 24 episodes.)	2,502,254
Japanese (L3) dramas	One Litre no Namida, One Piece, Imawa no Kuni no Alice (2 seasons), First love: Hatsukoi, Mystery to Inakare, Unnatural, 3 Nen A Gumi: Ima kara Mina-san wa, Hitojichi Desu, Hanzawa Naoki, Mother, Kazoku Game, Tenno no Ryoriban, MIU 404, Silent, Brush up Life, Boku no Ita Jikan, Tonbi, Soredemo, Ikite Yuku, Hanzawa Naoki, Ouroboros, Code Blue (3 seasons), Juhan Shuttai, Kyo Kara Ore wa, Liar Game, Koi wa Tsuzuku yo Doko Made mo, The Confidence Man JP, Boku Dake ga Inai Machi, Kono Sekai no Katasumi ni, Dele, Konto ga Hajimaru, Nanba MG5, 99.9: Keiji Senmon Bengoshi, Jin, N no Tame ni, Good Doctor, Doctor X, Nijjiro Karute, Bitter Blood, Bokura wa Kiseki de Dekite Iru, Cold Case: Shinjitsu no Tobira, Koi Desu: Yankee-kun to Hakujou Garu, Kounodori, Tengoku to Jigoku: Psychona Futari, Kieta Hatsukoi, Influence, Ore no Sukato, Doko Itta.	50 (Each series ranges from 8 to 12 episodes.)	2,490,044
American (L2) dramas	Young Sheldon (6 seasons), My Life with the Walter Boys, Sweet Home (2 seasons), the Crown (6 seasons), Bodies, Neon, Love is Blind (5 seasons), The Fall of the House of Usher, Lupin (3 seasons), Outlanders (7 seasons), Sex Education (4 seasons), Shameless (11 seasons), Painkiller, Alice in Borderland (2 seasons), Kleo, The Witcher (3 seasons), Insecure (5 seasons), Wednesday, Black Mirror (6 seasons), Suits (9 seasons), Never Have I ever (4 seasons).	80 (Each series ranges from 4 to 22 episodes.)	2,507,792
Chinese (L1) dramas	The Untamed, Nirvana in Fire, Hidden Love, Love like the Galaxy (2 seasons), Mysterious Lotus Casebook, Joy of Life, When I fly towards you, Go Ahead, The Blood of Youth, Meet Yourself, Love between Fairy and Devil, Eternal Love, The Story of Ming Lang, Story of Yanxi Palace.	15 (Each series ranges from 24 to 78 episodes.)	2,508,978

2.2 Instrument and Data Processing

Lexical text coverage refers to “the percentage of running words in the text known by the reader” (Nation, 2006, p. 61). Some researchers regard five unknown words in every 100 words (95% text coverage) as the boundary beneath which readers may not be able to get adequate comprehension of an authentic text (Laufer, 1989; Read, 2000). Nation (2006) noted that 98% text coverage, i.e. two unknown words per 100 words, is ideal for guessing words from context and may offer good conditions for lexical learning. These two putative coverage points have been proposed for setting vocabulary thresholds for different purposes: 95% text coverage for reasonable comprehension (Laufer & Ravenhorst-Kalovski, 2010) and 98% coverage for pleasure reading (Hirsh & Nation, 1992) as well as for unassisted reading (Hu & Nation, 2000).

As far as video viewing is concerned, a lower vocabulary threshold such as 95% lexical coverage may be a likely condition for guessing words from context successfully, since video affords viewers extra-linguistic support (e.g., body language) to gain understanding (van Zeeland & Schmitt, 2013). Visual clues may make it less difficult for EFL learners to guess meanings from context, even if their lexical knowledge falls short of 95% text coverage. In her recent research, Laufer (2020) detected that 95% text coverage could be reached with an initial knowledge of 90% words in the text plus inferring an additional 5% words. Webb and Rodgers (2009) even surmised that the necessary text coverage for adequate comprehension of television programs is likely to be as low as 90%.

A vocabulary threshold or level hinges on the predetermined text coverage point. For instance, 98% lexical coverage of Voice of America (VOA) news entails mastery of the first 6000 word families while 95% text coverage requires knowledge of the first 3000—4000 word families (Hsu, 2019). In line with past studies, this research adopted both 95% and 98% coverage as a cutoff for measuring the vocabulary levels of English subtitles. Through the word-frequency scale of the British National Corpus and the Corpus of Contemporary American English (BNC/COCA) ranking twenty-five 1000 words families by frequency and dispersion (Nation, 2016, 2017), the researcher counted how many of the ranked 1000-word-family lists from the first 1000 were needed until the total text coverage accumulated to 95% or 98%. Meanwhile, the vocabulary level was extrapolated according to which 1000-word-family list was the last one being added when the cumulative coverage reached 95% or 98%.

The AntWordProfiler Version 2.1.0 (Anthony, 2023) was implemented to analyze the lexical profiling of English subtitles. In addition to the ranked 25 BNC/COCA 1000-word-family lists, another four ever-growing lists of marginal words, proper nouns, abbreviations and transparent compounds compiled by Nation (2017) were also installed to the program. The words that were placed at the ‘Words NOT Found In Base Lists’ (hereafter called off-list) by the AntWordProfiler were further examined.

If off-list words were marginal words (ouch, huh, geez) and personal or geographical names, they were supplemented to the existing interjection list and proper noun list in turn. For hyphenated transparent compounds in the off-list, TextMate 2.0 was applied to replace these words with a space. The hyphens were summarily removed so that these compounds would not be mistaken by the AntWordProfiler for off-list words. Similarly, closed compounds in the off-list were added to the existing compound list to avoid double counting if their component words are already in the BNC/COCA word lists.

Proper nouns, spoken interjections, transparent compounds and abbreviations do not make reading arduous. Excluding their text coverage would inflate the vocabulary level and overestimate the vocabulary threshold required for adequate comprehension (Nation, 2006; Nurmukhamedov & Webb, 2019), since the reading load they put on is manageable.

It may not be difficult to recognize a proper noun due to transliteration or the capitalization of its first letter. Spoken interjections can be ignored because of little influence on reading comprehension. One can infer the meaning of a transparent compound from its constituent words effortlessly if the learner is already familiar with its individual words. Being infrequent, abbreviations do not pose a hurdle in subtitle reading. The meaning of an abbreviation may be talked about in subsequent dialogues, or it may be a familiar one in daily life. Following previous research taking the text coverage of these four types of words into account (Nation, 2006; Rodgers & Webb, 2011; Webb &

Rodgers, 2009), the researcher also counted them in to avoid overestimation until 95% and 98% coverage points were achieved.

2.3 Word Family as a Counting Unit for lexical coverage

This research involves reading English subtitles while viewing streaming TV series in different languages. For reading purposes, English word recognition is essential for comprehension. Word lemmas as a counting unit may be less preferable to word families in terms of an overestimate of the recognition vocabulary amount needed for good comprehension according to Webb (2021). A word lemma contains a base form and inflectional forms that belong to the same part of speech, while a word family additionally includes derivational affixes (Nation & Meara, 2010). For instance, *fuel* and *fuels* as nouns as well as *fuel*, *fuels*, *fuelled* and *fueling* in verb forms would be different lemmas. They would be categorized into two lemmas or counted as one word family, leading to different vocabulary sizes. To avoid overestimation, ‘word family’ was used as a counting unit for lexical coverage.

2.4 The New Vocabulary Levels Test and Interview

Ten college freshmen in the researcher’s class volunteered to watch English subtitled dramas of their choice in their spare time for one semester. They were given free access to Netflix for one year as a reward but they were asked to watch Korean, Japanese, Chinese, American dramas at least one series for each language within 4 months. Before viewing, the Vocabulary Levels Test (Webb, Sasao & Balance, 2017) was used to measure their vocabulary size. After four month’s viewing, they were interviewed regarding their perception of viewing English-subtitled TV series, especially their views on non-English-speaking dramas.

3. Findings and Discussion

3.1 Vocabulary Levels of English-subtitled Korean, Japanese, Chinese Dramas as opposed to American Dramas

Table 2 provides a snapshot of the vocabulary levels of English subtitles across TV series in four different languages at 95% and 98% text coverage and the coverage distribution among the BNC/COCA 1st to 25th 1000-word-family levels.

Table 2. Vocabulary Levels of English Subtitles Along the BNC/COCA Word-frequency Scale

Word list	English subtitles of Korean (L3) dramas		English subtitles of Japanese (L3) dramas		English captions of American (L2) dramas		English subtitles of Chinese (L1) dramas	
	Text coverage	Cumulative coverage	Text coverage	Cumulative coverage	Text coverage	Cumulative coverage	Text coverage	Cumulative coverage
<i>Proper nouns etc.</i>	3.27%	3.27%	3.20%	3.20%	3.08%	3.08%	3.33%	3.33%
1 st 1000	87.37%	90.64%	87.74%	90.94%	84.03%	87.11%	87.43%	90.76%
2 nd 1000	4.32%	94.96%	4.06%	*95.00%	5.01%	92.12%	4.12%	94.88%
3 rd 1000	1.98%	*96.94%	1.79%	96.79%	1.98%	94.10%	1.84%	*96.72%
4 th 1000	0.96%	97.90%	1.22%	**98.01%	1.65%	*95.75%	0.92%	97.64%
5 th 1000	0.83%	**98.73%	0.77%	98.78%	0.95%	96.70%	0.78%	**98.42%
6 th 1000	0.31%	99.04%	0.38%	99.16%	0.76%	97.46%	0.53%	98.95%
7 th 1000	0.21%	99.25%	0.08%	99.24%	0.52%	97.98%	0.44%	99.39%
8 th 1000	0.09%	99.34%	0.07%	99.31%	0.44%	**98.42%	0.27%	99.66%
9 th 1000	0.07%	99.41%	0.07%	99.38%	0.42%	98.84%	0.12%	99.78%
10 th —25 th 1000	0.59%	100%	0.62%	100%	1.16%	100%	0.22%	100%
Total words	2,502,254		2,490,044		2,507,792		2,508,978	

As shown in Table 2, the first 1000 word families accounted for 84.03%—87.74% of the total words in the sub-corpora and the second 1000 made up 4.06%—5.01%. The combined text coverage of the first 2000 word families (89.04%—91.8%) was higher than the average 86%—90% covered by

the first BNC 2000 word families in texts of different genres (Nation, 2006). This result offers a beacon of hope for EFL students with a base vocabulary only (at the 2000—3000 word-family levels).

A further look at Table 2 shows that after the first 2000 words, the subtitle coverage of the third 1000 reduced to below 2% and further dropped to less than 1% at the fifth 1000. English subtitles in non-English-speaking TV series used a small vocabulary, converging at the first 2000—3000 word-family levels. Knowledge of the first 2000—3000 word families plus interjections, proper nouns, transparent compounds and abbreviations would be enough to provide 95% coverage of English subtitles for reasonable or minimally acceptable comprehension (see Table 2 for cumulative coverage 96.94%, 95%, 96.72% at the 2nd and 3rd 1000 levels for Korean, Japanese and Chinese dramas respectively). Learners with knowledge of the first 4000—5000 word families (providing 98.73%, 98.01%, 98.42% coverage for Korean, Japanese and Chinese dramas) should be able to read English subtitles with less difficulty. In other words, watching English-subtitled non-English-speaking TV series like Korean, Japanese and Chinese dramas did not demand a higher vocabulary capacity than the others.

In contrast with non-English-speaking dramas, watching American TV series was the most vocabulary-demanding (3000—4000 word families at 95% coverage and stretching to 7000—8000 word families at 98% coverage). English subtitles at a higher vocabulary level can be interpreted as having more English words appearing in the latter 1000-word-family bands along the BNC/COCA word-frequency scale. Reaching the 8000 word-family level means that American series entail a much larger vocabulary than the other three non-English-speaking dramas and watching American series would result in EFL learners working on a wider variety of vocabulary subject to genres. However, the slightly below 95% coverage in American series that knowledge of the first 2000—3000 word families plus marginal words and proper nouns would provide may still fall within the parameters where good comprehension takes place (Nation, 2006).

Regardless of non-English-speaking or English-speaking dramas, if EFL learners have a vocabulary capacity of the first 8000 word families, they would be able to view English-subtitled TV series rather smoothly without constantly guessing unknown words or frequent dictionary lookups. Compared with a vocabulary size of the first 3000—4000 to 7000—8000 word families needed to watch American series, watching non-English-speaking dramas with English subtitles is a lot easier, since knowledge of the first 2000—3000 to 4000—5000 word families would suffice to provide 95% to 98% text coverage. However, this signals that watching non-English-speaking dramas may not bring about greater vocabulary development as watching American series.

In answer to RQ1 ‘What are the vocabulary levels of English-subtitled Korean, Japanese and Chinese dramas as opposed to American dramas?’, Table 2 shows that they generally reached the 2nd–3rd 1000-word-family levels at 95% text coverage and spread to the 4th–5th 1000 at 98% coverage as opposed to American dramas spanning from the 3rd–4th 1000 word-family level at 95% coverage to the 7th–8th 1000 level at 98% coverage. By and large, EFL drama buffs with knowledge of the first 2000 to 3000 word families can be encouraged to give viewing English-subtitled dramas a try, since they are supported with rich visual imagery, which further lowers the vocabulary demand. Therefore, watching English-subtitled dramas may not be a formidable task as we have expected. EFL learners planning to watch captioned American dramas may need to know a minimum of the first 3000—4000 word families in order to enjoy binge-watching American series.

As aforementioned, the lowest levels of graded readers contain only a few hundred headwords, most go up to around 3000 headwords, and the very highest levels go up to 4000—5000 headwords, equivalent to the first 3000—4000 word-family levels at 98% text coverage. Compared with graded readers, the vocabulary levels of English subtitles in TV series are above the most advanced graded reader level. Moreover, TV dramas often involve some societal issues to which global audience can relate to, such as bullying, corruption, gender inequality, sexual harassment as well as daily life topics like family, friendship and marriage. The diversity of subject matter in TV dramas implies that English subtitles thereof may be a rich language learning resource.

3.2 Learners’ perception towards viewing English-subtitled dramas

Based on the Vocabulary Levels Test (Webb, Sasao & Balance, 2017) scores, the recognition vocabulary of the ten participants ranged from the first 2000 to 6000 word-family level. Seven of them converged at the 3000—4000 word-family levels with one student reaching the 6000 word-family

level and two with a base vocabulary at the 2000—3000 word-family level. At the end of the semester, all of the ten participants had completed the English-subtitled drama viewing assignment—at least one Korean, Japanese, Chinese and American series of their choice respectively.

Subsequently, they were interviewed to express their views on watching English-subtitled Korean and Japanese (L3), American English (L2) and Chinese (L1) TV series. The ten participants showed a level of agreement in their responses regarding the difficulty of English subtitles. They all commented that viewing English-subtitled Korean, Japanese and Chinese dramas was not so formidable as they had expected, but they felt a little more difficult while watching American series. This may be because the vocabulary sizes of the participants have all attained a minimum threshold at the first 2000 to 3000 word-family levels for Korean, Japanese, Chinese dramas but the vocabulary capacities of two participants fell below the minimal demand of the first 3000 to 4000 word-family levels for American series. Their responses concurred with some lexical researchers' proposition that 95% text coverage (percentage of known words in a text) is ideal for reasonable comprehension. Or it was likely that through visual representations and cumulative background knowledge over episodes and series, learners with knowledge of words covering slightly lower than 95% of subtitled text may still provide general comprehension of the scenarios and plots.

Not very surprisingly, most of them admitted that they did not pay much attention to English subtitles when they watched Chinese dramas, since they could hear Mandarin (L1) dialogues effortlessly. For this issue, one of them jokingly suggested turning off the audio or turning down the volume. As to Korean and Japanese dramas, they remarked that they had no choice but to rely on English subtitles since they knew little about Korean or Japanese. One participant even used the idiom 'kill two birds with one stone' to tell the researcher that he can learn English together with spoken Korean. Then he blurted out 매우 감사합니다, meaning 'thank you very much'.

When asked about their preference for English-speaking dramas or non-English-speaking dramas, most of them answered that a gripping plot instead of the language spoken is the key to keeping them hooked and binge-watching a TV series. Although only one student showed her preference for American dramas and her English was the best among the ten participants, she still felt that watching American series was more arduous than Korean, Japanese or Chinese dramas to understand the characters' dialogues per scene. To overcome the hindrance from unknown words, this student said that she used a guessing strategy by picking up key words from a sequence of utterances and subtitles for general comprehension. In this regard, the two low-proficiency students (having a base vocabulary only) mentioned that they could guess meanings without difficulty when they watched Korean or Japanese dramas. They gave their reason that the mode of thought and culture in Korean dramas or Japanese dramas are close to those in Chinese dramas.

Some other participants also seconded the two students' viewpoints about easier guess due to similar culture. Very surprisingly, one of them pointed out the word 'waste' as an example. This participant described a scene she saw in a Korean drama, where a man gives a piggyback to his drunk female friend, who just broke up with her boyfriend, and climbs stairs to the rooftop hut of an old mid-rise apartment without elevators, saying "Shush! You are wasted. Do you want to wake everyone to know you were dumped?" This participant said that she already knew the usual meaning of 'waste', signifying use or expend carelessly, but 'You are wasted' impressed her and thus learned a new meaning of 'waste' from this English subtitle. She further explained that neither 'You are useless' nor 'You are a jerk' was what she inferred from the scene. Instead, she guessed the meaning 'You are drunk' correctly, because such a dramatic scene occurs occasionally in Chinese dramas.

Overall, the interview reveals that regardless of English-speaking dramas or non-English-speaking dramas, learners would still likely gain from English subtitles. They would likely uptake the words that are related to their life experiences and relatable scenes may motivate them to use some strategies to learn new words or develop further knowledge of familiar words.

4. Conclusion and Implications

This research was inspired by the binge-watching phenomenon among Taiwanese college students. Because our students enjoy watching dramas on OTT in their leisure time and spend so much time binge-watching, this hints that drama fever may make it possible for EFL learners like ours to continually engage in watching dramas with English subtitles in lieu of their L1 subtitles. The goal of

this preliminary study has been to show the potential of English subtitles as an alternative to graded readers for extensive reading.

English-subtitled Korean, Japanese and Chinese dramas reached the 2nd—3rd 1000 word-family levels at 95% text coverage and the 4th—5th 1000 levels at 98% coverage subject to genres, while American series extended to the 7th—8th 1000 level at 98% coverage from the 3rd—4th 1000 word-family levels at 95% coverage. Learners with a base vocabulary may find English subtitles a challenge at the beginning. However, the interview results tell another story. For the ten participants, incomplete comprehension was expected when watching the first few episodes, but they did not feel overwhelmed by the amount of unknown words. Visual imagery assisted them in creating the form-meaning link of unknown words and therefore made up for insufficient lexical knowledge.

To overcome fear of watching English-subtitled series at the onset, learners can be advised to proceed to watch further on. Comprehension will improve with more and more episodes being watched, because the context of an unfamiliar word and its later recurrence may give viewers some clues about its meaning. There are still some other ways of helping EFL learners to reduce vocabulary load. Gaining background knowledge through trailers and preview facilitates comprehension. With the subtitle option on OTT, EFL learners can select the L1 subtitles in the first time viewing and watch the same drama later but in English subtitles. Therefore, learners may find it less difficult in the second time viewing. Another way to increase the potential for vocabulary learning is to watch the same English-subtitled drama multiple times (at spaced intervals) if they enjoy this drama. This would increase the number of encounters with unfamiliar words to the point where they may be learned.

As far as the vocabulary levels involved in English-subtitled Korean, Japanese and Chinese dramas are concerned, this research has manifested that EFL learners can get sufficient input from English subtitles through binge-watching, just as they do extensive reading of graded readers. The results may serve as a reference for English teachers and learners who are concerned with vocabulary learning out of class.

Acknowledgment

The author wishes to express her profound gratitude to Emeritus Professor Paul Nation for the BNC/COCA word lists, which were indispensable to this research.

Declarations

- Author contribution** : Wenhua Hsu is the sole author and responsible for all of this manuscript.
- Funding statement** : This research was self-supported and did not rely on any institutional funding.
- Conflict of interest** : The author declares no conflict of interest.
- Ethics declaration** : The author acknowledges that this work has been written based on ethical research that conforms with the regulations of the author's university and that she has obtained the participants' permission when collecting data. The author supports English Language Teaching Educational Journal (ELTEJ) in maintaining high standards of personal conduct, practicing honesty in all our professional practices and endeavors.
- Additional information** : No additional information is available for this paper.

REFERENCES

- Anthony, L. (2023). *AntWordProfiler* (Version 2.1.0) [Computer software]. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software>
- Claridge, G. (2012). Graded readers: How the publishers make the grade. *Reading in a Foreign Language*, 24(1), 106–119. <http://hdl.handle.net/10125/66668>

- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge University Press.
- Day, R. R., & Bamford, J. (2004). *Extensive reading activities for teaching language*. Cambridge University Press.
- Dizon, G. (2016). Online video streaming in the L2 classroom: Japanese students' opinions towards Netflix and subtitles. *Osaka JALT Journal*, 3, 70-87. <https://doi.org/10.4995/eurokal.2018.9080z>
- Dizon, G. (2018). Netflix and L2 learning: a case study. *The EuroCALL Review*, 26(2), 30-40. <https://doi.org/10.4995/eurocall.2018.9080>
- 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. <http://hdl.handle.net/10125/67046>
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61(3), 355-382. <https://doi.org/10.1353/cml.2005.0018>
- Hsu, W. (2019). Voice of America (VOA) news as voluminous reading material for mid frequency vocabulary learning. *RELC Journal*, 50(3), 408-421. <https://doi.org/10.1177/0033688218764460>
- 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>
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From human thinking to thinking machines* (pp. 316-323). Multilingual Matters.
- Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, 54(4), 1076-1085. <https://doi.org/10.1002/tesq.3004>
- 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. <http://hdl.handle.net/10125/66648>
- Locher, M. A., & Messerli, T. C. (2020). Translating the other: Communal TV watching of Korean TV drama. *Journal of Pragmatics*, 170(1), 20-36. <https://doi.org/10.1016/j.pragma.2020.07.002>
- Markham, P., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems*, 31(3), 331-341. <https://doi.org/10.2190/BHUH-420B-FE23-ALA0>
- McQuillan, J. (2016). What can readers read after graded readers. *Reading in a Foreign Language*, 28(1), 63-78. <https://doi.org/10125/66715>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, 27(1), 136-145. <http://hdl.handle.net/10125/66705>
- 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 set]. <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>

- Nation, I. S. P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd Edition) (pp. 34–52). Edward Arnold.
- Netflix. (2014). *Netflix declares binge watching is the new normal*. <https://www.prnewswire.com/news-releases/netflix-declares-binge-watching-is-the-new-normal-235713431.html>
- Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, 52(2), 188-200. <https://doi.org/10.1017/S0261444819000028>
- Pedersen, J. (2019). Fansubbing in subtitling land: An investigation into the nature of fansubs in Sweden. *Target*, 31(1), 50–76. <https://doi.org/10.1075/target.18017.ped>
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Reynolds, B.L., Cui, Y., Kao, C. W., & Thomas, N. (2022). Vocabulary acquisition through viewing captioned and subtitled video: A scoping review and meta-analysis. *Systems*, 10, 133-152. <https://doi.org/10.3390/systems10050133>
- 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>
- Rubenking, B., & Bracken, C. C. (2021). Binge watching and serial viewing: Comparing new media viewing habits in 2015 and 2020. *Addictive Behaviors Reports*, 14, 100356. <https://doi.org/10.1016/j.abrep.2021.100356>
- 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. <http://dx.doi.org/10.1093/applin/ams074>
- Waring, R., & McLean, S. (2015). Exploration of the core and variable dimensions of extensive reading research and pedagogy. *Reading in a Foreign Language*, 27(1), 160–167. <http://hdl.handle.net/10125/66708>
- Webb, S. (2021). Word families and lemmas, not a real dilemma: Investigating lexical units. *Studies in Second Language Acquisition*, 43(5), 973–984. <https://doi.org/10.1017/S0272263121000760>
- Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366. <https://doi.org/10.1111/j.1467-9922.2009.00509.x>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics*, 168(1), 34–70. <https://doi.org/10.1075/itl.168.1.02web>