Journal of Education and Learning; Vol. 13, No. 6; 2024
ISSN 1927-5250 E-ISSN 1927-5269
Published by Canadian Center of Science and Education

Reversed Subtitling and Extensive Reading: The Case of English-Subtitled Mandarin Dramas

Wenhua Hsu[1]

[1] International College, I-Sou University, Kaohsiung, Taiwan

Correspondence: Wenhua Hsu, International College, I-Sou University, Kaohsiung, Taiwan.

Received: April 6, 2024 Accepted: June 9, 2024 Online Published: July 26, 2024

doi:10.5539/jel.v13n6p1 URL: https://doi.org/10.5539/jel.v13n6p1

Abstract

The binge-watching phenomenon on college campuses in Taiwan inspired this study. The researcher often overhears her students chatting about which Mandarin TV series they have been binge-watching recently. Given this drama fever, which may provide an impetus for sustained reading of on-screen text, the researcher is concerned with English vocabulary growth if the viewing habit shifts from Mandarin to English subtitles. A corpus of over 5.6 million English-subtitled words from 37 Mandarin dramas was compiled, totaling 1,238 episodes. The operational measures involved the ranked twenty-five 1000-word-family lists along the British National Corpus and the Corpus of Contemporary American English word-frequency scale. Results show that Mandarin drama English subtitles reached the 2000–3000 word-family levels at 95% text coverage and extended to the 4000–5000 levels at 98% coverage subject to genres. EFL Mandarin drama fans may encounter most words from each of the 1st to 6th 1000-word-family lists twelve times or more for potential learning by continually watching up to 24 English-subtitled Mandarin dramas. Moreover, twenty participants expressed their views on watching English-subtitled Mandarin dramas to a certain level of agreement. For extensive reading practitioners, the results may be a reference concerning what vocabulary level EFL learners may attain if they binge-watch English-subtitled Mandarin dramas in their leisure time.

Keywords: reversed subtitling, lexical text coverage, vocabulary level, BNC/COCA

1. Introduction

As a result of the rapid rise of over-the-top (OTT) streaming platforms (e.g., Netflix, Hulu) since the late 2000s, the way people watch television (TV) series has changed a lot. Especially for Generation Y and Z people, there has been a shift in viewing modes from traditional appointment viewing to viewing on streaming platforms. Appointment viewing means sitting on the couch in front of a television set to watch TV programs in a once-daily or once-weekly broadcasting format. At the same time, videos on streaming platforms through mobile devices with Internet connectivity can be accessed anywhere, anytime. The latter mode has led to binge-watching with more ease. As per the Cambridge Dictionary, 'binge-watch' is to watch several episodes of a TV series or program one after another. This new viewing behavior has been defined by Netflix (2014) as viewing from two to six episodes in one sitting based on nearly 1500 TV streamers in the survey.

Streaming services have also brought in the phenomenon of binge-watching on college campuses in Taiwan, where English as a foreign language (EFL) is a required course for non-English-majoring students. In College English class, the researcher-teacher often overhears her students chatting about which drama series they have been binge-watching lately. Binge-watching series has become one of the leisure activities in their free time.

On OTT platforms, which aim at a global audience, English is the primary language into which video content is 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). Nevertheless, it is unsurprising that EFL Taiwanese students choose Mandarin subtitles (the first language, L1) when they watch Mandarin dramas. Given this, this research concerns English vocabulary growth if the students can change their viewing habits from Mandarin to English subtitles.

Subtitles are on-screen text in the viewer's first language (L1) as opposed to L2 captions in L2 videos (Markham & Peter, 2003). Different from the two normal situations (L1 subtitles and L2 captions in L2 videos), this research aimed at reversed subtitling (Zanon, 2006, 2007), where video content is delivered in the L1 and subtitles are in

the L2 (i.e., English, the target language). Suppose we draw a parallel between voluminous reading of English subtitles while binge-watching drama series and extensive reading of graded readers. In that case, they share common purposes, such as reading for pleasure, fluency, and vocabulary growth (Day & Bamford, 2004). As such, English subtitles may serve as an alternative to graded readers, which college English teachers like ours often encourage EFL students to read in an extensive reading scheme in their first year. However, compared with the time spent reading graded readers, the time that our students spend binge-watching in their spare time is a lot more. For this reason, this research addressed three questions concerning English subtitles as a source of input and the vocabulary learning opportunities they afford.

1) What are the vocabulary levels of English subtitles in Mandarin dramas?

2) By watching English-subtitled Mandarin dramas for one to three years during college, how many words beyond the first 3000 word families may EFL learners encounter often enough for vocabulary learning?

3) What are the learners' perceptions towards viewing English-subtitled Mandarin dramas?

2. Literature Review

2.1 Past Studies on Subtitling

Subtitles are screen text representing audio content in a film, television show, opera, or other audiovisual media ("Subtitles," 2024). They are lines of text at the bottom of the screen that translate spoken dialogues into the audience's language, which differs from that used in the source video. In accordance with previous research on subtitling and its impact on language learning, Talaván (2006) summarized subtitling modalities into three kinds: bimodal subtitles, standard subtitles, and reversed subtitles. Bimodal subtitles are on-screen text when the audio and subtitles are in the same foreign language, called L2 captions. Standard subtitles are on-screen text in the viewer's mother tongue while the audio is in a foreign language. In contrast, reversed subtitling (L1 audio with subtitles in L2) refers to text when the audio is in the mother tongue while subtitles are in a foreign language.

Through a scoping review and meta-analysis, Reynolds, Cui, Kao and Thomas (2022) scrutinized 139 past studies to explore how viewing L2-captioned and L1-subtitled videos can lead to effective vocabulary learning. They concluded that captioned or subtitled videos positively affected vocabulary learning. Reversed subtitling offers benefits, especially to less advanced learners, in vocabulary learning and grammar acquisition (Zanon, 2006, 2007). In a pioneer study on reversed subtitling, Lambert, Boehler, and Sidoti (1981) discovered that for Grades 5–6 students in a French immersion program, reversed subtitling was the most beneficial mode regarding L2 comprehension, contextual meaning, and spelling. They explained that the L1 auditory input facilitated L2 comprehension and thus provided time for effective L2 text processing. In a follow-up test on whether the positive effect of the subtitling intervention can still be maintained after long-term exposure (Holobow, Lambert, & Sayegh, 1984), the results demonstrated that the group with sustained reversed subtitling performed best in contextual meaning and comprehension tests. They concluded that L2 beginners may face difficulties in L2 processing when the input is provided solely in the target language. In contrast, reversed subtitling provided a favorable condition where L1 was present in at least one audiovisual channel to assist the learners in L2 comprehension.

In a similar line of research, Danan (1992) compared standard and reversed subtitling based on the dual coding theory (Paivio, 1971, 1986). She offered that the participants received comprehensible L1 input by watching reversed subtitled video, enabling them to engage in deeper processing of L2 text. Danan (1992) stated that a reversed subtitled video provides a triple connection between image, sound in one language, and text in another. The instant association of a foreign word with its L1 referent leads to incidental vocabulary learning. In later research, Danan (2004) further advocated reversed subtitling for better language processing and recall because the presence of images, sound in the native language, and text in the target language helps to retrieve information from visual and verbal traces.

To examine which subtitling mode achieves the best result concerning comprehension and vocabulary recall, Bairstow and Lavaur (2011) used L2 captioning, L1 subtitling, and reversed subtitling for comparison in their research. While the comprehension scores under the captioning mode were similar to those under the subtitling mode, the average scores on the comprehension and vocabulary recall tests were the highest under the reversed subtitling mode. In line with Bairstow and Lavaur (2011), Fazilatfar, Ghorbani, and Samavarchi (2011) compared the effects of subtitling, reversed subtitling, and no subtitling on vocabulary learning. The results showed that the reversed subtitling mode positively affected word form recognition, meaning translation, and passive/receptive vocabulary learning. As to learners' perceptions regarding reversed subtitling, Čepon (2011) reported that all the university-level participants responded positively to this method and presumed that reversed subtitling was beneficial in vocabulary acquisition. The quantitative results also supported reversed subtitling in this regard.

Gorijan (2014) investigated the effect of film subtitling on incidental vocabulary learning by employing bimodal subtitling, standard subtitling, and reversed subtitling among three groups of Iranian English-majoring students, totaling 90 participants. Before playing the animated movies with three treatments, they were given a vocabulary pre-test with forty multiple-choice questions and a post-test after the subtitling intervention. The results revealed that reversed subtitling was the most effective mode of subtitling in incidental vocabulary learning, followed by bimodal subtitling and then standard subtitling. However, in recent research by Kwarteng (2023), standard subtitling (subtitles in L1) has a more significant effect on incidental vocabulary learning than bimodal subtitling (subtitles in L2).

Another study by Kim (2020) concerned the effects of subtitling on EFL learners' receptive skills. One hundred and eight Korean EFL learners took TOEIC listening and reading tests after exposure to subtitled materials for ten weeks. The TOEIC scores showed that the group receiving reversed subtitling improved English reading skills most, while the group receiving standard subtitling enhanced English listening abilities most.

The literature review above provides some implications for this research regarding vocabulary learning in reversed subtitling (L1 videos with L2 subtitles). Namely, video in the native language may help learners understand subtitles better in the target language for subsequent language processing and recall.

2.2 Extensive Reading

Informed by the input hypothesis within second language acquisition (Krashen, 1985), extensive reading has long been promoted to complement intensive reading for language learning (Day & Bamford, 1998, 2004; Horst, 2005). Learners are encouraged to read materials of interest that are lower than or equivalent to their proficiency level extensively to avoid looking up words in a dictionary frequently (Waring & McLean, 2015). The rationale behind extensive reading is that learners need to engage in reading in amounts large enough to encounter unfamiliar or unknown words sufficiently to internalize how those words are used and how they collocate with neighboring words (Nation, 2015).

To carry out an extensive reading scheme, college English teachers in the current EFL context often use grade reader series, which are books of simplified fiction and non-fiction classics with controlled vocabulary and grammar, to make them accessible to intermediate-level EFL learners. There are 250 to 3800 headwords from Level One to Level Six or Seven in a grade reader series (Claridge, 2012). The highest levels from Cambridge Readers and Oxford Bookworms reach the first 4000 word-family level (McQuillan, 2016).

The present situation is relevant to extensive reading when Taiwanese college students use English (L2) subtitles instead of Mandarin (L1) subtitles while watching their favorite Mandarin TV series. As with graded readers, English subtitles in this scenario are lexical learning resources of value. Accordingly, the vocabulary levels of English subtitles in Mandarin dramas are an essential issue that needs to be examined first.

2.3 Vocabulary Levels of TV Programs and Narrow Viewing

Webb and Rodgers (2009) collected eighty-eight English scripts of American and British TV series across various genres and measured the vocabulary levels thereof along the word-frequency scale of the British National Corpus, which ranks words from the first to the fourteenth 1000-word-family levels based on frequency and dispersion. Using 95% and 98% lexical text coverage (95% and 98% of the total words in a text known) as the lower and upper thresholds, they gauged 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 (i.e., five unknown words every 100 words). In contrast, 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 effectively using TV programs for English learning, Rodgers and Webb (2011) recommended narrow viewing by watching TV programs with related themes rather than random episodes from programs with different themes or watching a series of programs with similar themes or storylines. They offered some evidence that the number of words occurring once was more significant, 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 unrelated TV programs may use a wider variety of words and are likely to have a higher vocabulary load than programs with similar themes or a story arc linking episodes. Related programs may use fewer words but with higher frequency than unrelated programs, yet for learners, the chance of having a more significant number of encounters with those words is higher, increasing the likelihood of vocabulary learning.

It is worth highlighting that the current corpora comprise English subtitles of Mandarin-speaking TV series instead of English scripts of English-speaking TV programs as in Webb and Rodgers' (2009) corpora. Undeniably, it is more advisable for EFL learners to watch English-speaking dramas with or without English captions than non-

English-speaking dramas but with English subtitles. However, EFL learners with Mandarin drama fever can marathon a Mandarin TV series in one sitting during holidays or engage in "serial viewing" (Rubenking & Bracken, 2021, p. 100356) over several days or weeks in their spare time. Therefore, the researcher argues that this impetus allows avid Mandarin drama fans to sustain reading English subtitles while binge-viewing Mandarin dramas on OTT platforms.

2.4 Lexical Text Coverage for Measuring Vocabulary Levels

Lexical text coverage refers to "the percentage of running words in the text known by the reader" (Nation, 2006, p. 61). The principle behind lexical text coverage is how many unknown words in a text can be tolerated before they interfere with comprehension. Some researchers consider five unknown words in every 100 words (95% lexical text coverage) as the boundary beneath which readers may be unable to get adequate comprehension of an authentic text (Laufer, 1989; Read, 2000). A lack of familiarity with more than 5% of the running words in a text may make reading an arduous task for EFL learners in terms of reading fluency.

Nation (2006) presumed that for adequate comprehension, 98% lexical text coverage, i.e., two unknown words per 100 words, is ideal for guessing words from context and may offer good conditions for incidental vocabulary learning. Nevertheless, lexical text coverage in percentage is not equal to the degree of comprehension. Researchers have diverged in the percentage of lexical text coverage necessary for adequate comprehension. In the literature, two putative coverage points have been generally adopted for setting vocabulary thresholds for different purposes: 95% lexical 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 Laufer (1989) has pointed out, the two coverage points imply probabilistic boundaries over which learners can gain comprehension to a certain extent.

As to video viewing, a lower vocabulary threshold (less than 95% coverage) may be a likely condition for successfully guessing from context. Video affords viewers extra linguistic support (e.g., body language) to help comprehension (van Zeeland & Schmitt, 2013). Visual clues may make it less complicated for EFL learners to guess meanings from context, even though their lexical knowledge falls short of 95% text coverage. In her recent research, Laufer (2020) detected that 95% text coverage could be achieved with initial knowledge of 90% of the text plus inferring an additional 5% of words. To some degree, guessing correctly lies in the amount of contextual information to help infer meanings (Webb, 2008). Webb and Rodgers (2009) surmised that the necessary text coverage for adequate comprehension of television programs is likely to be as low as 90%. It is worth noting that different from Webb and Rodgers' (2009) research involving intra-lingual text (i.e., English subtitles in English-speaking TV programs), in the current situation, learners primarily receive comprehensible Mandarin (L1) audio while reading on-screen English (L2) text. Thus, there is some reason to believe that the instant mapping of L2 lexical form to L1 meaning leads to a lower lexical threshold necessary for adequate comprehension. However, in line with previous studies, this research also adopted 95% and 98% coverage as a cutoff for measuring the vocabulary levels of English-subtitled Mandarin dramas.

2.5 The Number of Exposure to a New Word for Incidental Learning

Research on incidental learning has documented that a single exposure to a new word rarely sustains robust learning of it (Horst, 2013; Horst, Parsons, & Bryan, 2011). Referring to previous studies on the number of encounters needed for word learning, ranging from 5 to 16 times, Nation (2014) reasoned that for uptake to happen, a word's occurrence in a variety of contexts totaling 12 times would just be enough to develop knowledge of that word. Following Nation (2014), the researcher adopted 12 occurrences as a cutoff to measure how many words beyond the first 3000 word families EFL learners may encounter often enough for vocabulary learning.

3. Method

3.1 The Corpus

The corpus contained 5,646,773 English subtitled words from 37 Mandarin TV series across several genres, available globally on OTT platforms (see Table 1). The 37 Mandarin TV series are among the highest-viewed dramas in the Mandarin-speaking regions. Although a TV series may be categorized into a specific drama genre (e.g., melodramas, comedy, romance, time travel, fantasy, coming of age, action, crimes, thriller, pop idol dramas, costume drama, historical, legal, medical and motivational dramas), the scenarios or the storylines often involve some societal issues, such as corruption, sexual harassment, wealth/gender inequality, and bullying as well as a couple of common topics like family, marriage, friendship, morality, (political) tactics and trickery, and conflicts. The diversity of subject matter in Mandarin TV series suggests that their English translations/subtitles may be a rich English learning resource, presenting close to real-life language use in reflecting culture and thinking.

The inclusion of popular Mandarin dramas was based upon the assumption that high viewership on OTT platforms is likely to have been watched by a large number of Mandarin drama fans around the world. It is worth highlighting that most Mandarin 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 crime, romance, thriller, or even comedy, leading to many twists and turns, say, along with homicidal scenarios and legal interrogations. In terms of corpus size, genre balance, and high viewership, the present corpus should suffice to provide a reliable measurement of vocabulary levels thereof.

Table 1. The Mandarin drama English subtitle corpus

Drama title	Genres	Number of episodes	Number of words
Some Day or One Day (Want to See You)	Time travel, romance	13 (90 min.)	135,231
In Time with You (I Might Not Love You)	Romance	13 (80 min.)	127,504
The Distance Between Us and Evil	Social realism	10 (50 min.)	81,257
Wave Makers	Political	8 (44–56 min.)	68,024
Copycat Killer	Crime, thriller	10 (52–71 min.)	80,453
Oh No! Here Comes Trouble	Comedy, fantasy	12 (52 min.)	82,456
Port of Lies	Legal	8 (50–56 min.)	69,377
Light the Night (3 seasons)	Thriller, mystery	72 (45–55 min.)	400,103
Mad doctor	Medical	10 (60 min.)	81,004
The Victim's Game	Crime, thriller	8 (62–68 min.)	70,448
The Making of an Ordinary Woman (2 seasons)	Comedy	40 (46–58 min.)	241,886
U Motherbaker	melodrama	40 (60 min.)	252,172
The Tricks of Boys and Girls	Pop idol drama	15 (60 min.)	1114,23
It Started with a Kiss (2 seasons)	Romance	40 (70–90 min.)	297,831
The Knockout	Crime	39 (45 min.)	225,332
Lost You Forever	Mythology	39 (43–51 min.)	230,118
Meet Yourself	Romance	40 (43–49 min.)	228,003
The Untamed	Costume drama	50 (45 min.)	280,009
Nirvana in Fire	Historical, politics	54 (40–45 min.)	300,966
Hidden Love	Youth, romance	25 (45 min.)	142,250
My Journey to You	Spy, romance, costume	24 (60 min.)	162,222
Imperfect Victim	Legal, crime	29 (45 min.)	175,032
Till the End of the Moon	Fantasy, time travel	40 (55 min.)	230,997
Fireworks of My Heart	Romance	44 (42 min.)	240,044
Blossoms Shanghai	Period drama	30 (46 min.)	166,777
Our Interpreter	Urban emotional drama	36 (45 min.)	210,908
Mysterious Lotus Casebook	Costume, martial arts	40 (45 min.)	222,334
Born to Run	Motivational drama	28 (45 min.)	162,774
Eternal Love	Fantasy, romance, costume	58 (45 min.)	314,473
Empresses in the Palace	Historical fiction	76 (45 min.)	411,116
Scarlet Heart	Historical fiction	40 (45 min.)	224,441
I Will Find You a Better Home	Workplace, motivational	53 (45 min.)	297,650
Meteor Garden	Idol drama	49 (41–47 min.)	252,954
A Journey to Love	Period legend, politics	40 (45–57 min.)	251,320
The Ordinary Glory	Workplace, urban	41 (45 min.)	255,861
The King's Avatar	Esports, crime, action	41 (45 min.)	218,108
A Love So Beautiful	Coming-of-age	23 (45 min.)	145,253
Total			5,646,773

Mandarin drama English subtitles in SubRip Text (SRT) files were downloaded from the Internet for research purposes. They are plain text files, including subtitles in sequence and the start and end time codes. 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 One of Some Day or One Day (alternatively titled, Want to See You), a blockbuster series on Netflix Taiwan in 2019–2020, later remade into the Korean drama A Time Called You, released in 2023.

51

00:07:43,790 --> 00:07:46,360

Everyone, imagine this.

52

00:07:46,360 --> 00:07:49,730

Right now, on this Earth,

53

00:07:49,730 --> 00:07:53,220

there's a person

who shares no blood relation with you

54

00:07:53,220 --> 00:07:56,330

but looks just like you.

55

00:07:56,330 --> 00:08:00,590

Like a version of you

from a parallel universe.

56

00:08:00,590 --> 00:08:04,630

You. You. You. You.

57

00:08:04,630 --> 00:08:07,960

Would you try finding them?

3.2 Instrument and Data Processing

A vocabulary demand or level hinges on the predetermined lexical text coverage point (% of words in a text known). For instance, 98% coverage of TED Talk English transcripts entails mastery of the first 5000—6000 word families, while 95% coverage requires knowledge of the first 3000 word families (Hsu, 2020). In the Literature Review, lexical researchers have generally adopted 95% and 98% coverage as lower and upper thresholds for good comprehension. Through the word-frequency scale of the British National Corpus and the Corpus of Contemporary American English (BNC/COCA), ranking twenty-five 1000-word families by frequency and dispersion (Nation, 2016), the researcher could size up the vocabulary levels of English subtitles in Mandarin dramas. The key to the measurement was to count how many ranked 1000-word-family lists from the first 1000 were needed until the total text coverage was 95% or 98%. Meanwhile, the vocabulary level was extrapolated, according to which 1000-word-family list was the last 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 complied by Nation (2017) were also installed in the program. The words 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. 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 the AntWordProfiler would not mistake these compounds 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.

Recognizing a proper noun may not be difficult due to transliteration or the capitalization of its first letter. Spoken interjections can be ignored because they have little influence on reading comprehension. One can effortlessly

infer the meaning of a transparent compound from its constituent words 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.

3.3 Word Family as a Counting Unit for lexical coverage

This research involves reading English subtitles while viewing streaming Mandarin TV series. 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 belonging to the same part of speech, while a word family also includes derivational affixes (Nation & Meara, 2010). For instance, *abandoning* and *abandonment* as nouns as well as *abandon*, *abandons*, *abandoned*, and *abandoning* in verb form 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.

3.4 The New Vocabulary Levels Test and Interview

Twenty college freshmen in the researcher's class volunteered to watch English-subtitled Mandarin 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 choose English subtitles while watching their favorite Mandarin TV series and to watch at least two dramas within four months. Before viewing, the Vocabulary Levels Test (Webb, Sasao, & Balance, 2017) was used to measure their vocabulary size. After four months of viewing, they were interviewed regarding their views on the reversed subtitling mode.

4. Results and Discussion

4.1 Vocabulary Levels of English-subtitled Mandarin Dramas

Table 2 provides a snapshot of the vocabulary levels of English subtitles in Mandarin dramas at 95% and 98% text coverage and the coverage distribution among the BNC/COCA 1st to 25th 1000-word-family levels. The over-5.6-million-token corpus comprised 37 Mandarin dramas plus three of them with sequels, totaling 1,238 episodes and involving 11,360 word families.

The first 1000 word families accounted for 87.43% of the total words in the corpus, and the second 1000 made up 4.12%. The combined text coverage of the first 2000 word families (=91.55%) was higher than the average 86%–90% covered by the first BNC 2000 word families in texts of different genres (Nation, 2006). This result (91.55%>86%–90%) offers a beacon of hope for EFL students with a base vocabulary only (at the 2000–3000 word-family levels).

Table 2. Coverage of the BNC/COCA word lists in English subtitles of 37 Mandarin dramas

Word lists	Tokens	Text coverage	Cumulative coverage	Word families
Proper nouns, etc.	188,038	3.33%	3.33%	X
1st 1000	4,936,974	87.43%	90.76%	1,000
2nd 1000	232,647	4.12%	94.88%	1,000
3rd 1000	103,901	1.84%	***96.72%**	995
4th 1000	51,950	0.92%	97.64%	991
5th 1000	44,045	0.78%	****98.42%**	939
6th 1000	29,928	0.53%	98.95%	887
7th 1000	24,846	0.44%	99.39%	801
8th 1000	15,246	0.27%	99.66%	746
9th 1000	6,776	0.12%	99.78%	696
10th–25th 1000	12,423	0.22%	100.00%	3,305
Total	5,646,773	100.00%		11,360

A further look at Table 2 shows that after the first 2000 word families, the subtitle coverage of the third 1000 was reduced to below 2% and dropped to less than 1% at the fourth 1000. This reveals that English subtitles in Mandarin dramas use 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 fluent reading (see Table 2 for cumulative coverage of 96.72% at the 3rd 1000 levels). If EFL learners have a vocabulary capacity of the first 5000 word families (providing 98.42% coverage), they can read English subtitles relatively smoothly in terms of the frequency of encountering unfamiliar words. In answer to RQ1 'What are the vocabulary levels of English subtitles in Mandarin 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. By and large, EFL Mandarin drama fans may need to know a minimum of the first 2000–3000 word families and, optimally, 4000–5000 word families to enjoy binge-watching English-subtitled Mandarin dramas.

4.2 Number of Words Beyond the First 3000 Word Families with Occurrence over Twelve Times

It may be feasible for Mandarin drama fans to watch one Mandarin drama series for one month and 12 dramas per year with a winter and extended summer vacation. A casual interview of Mandarin drama binge-viewers confirmed the presumption that they could view one TV series within one month on average. Assuming fans view 12 Mandarin dramas per year and 36 dramas over three years, it is equivalent to exposure to about 1.5 million English-subtitled words yearly and 4.5 million for three years (see Table 1 for 37 dramas totaling 5,646,773 words). It is worth adding that a conservative assessment was taken by calculating three years instead of four years. It was presumed that college students may be busy with their career plans in the fourth year before graduation and may no longer have time to binge-watch Mandarin dramas as before. To measure three different-sized datasets, a random selection of 12, 24, and 36 Mandarin dramas from the current corpus was repeated five times respectively, and the results of each dataset were averaged by five times of measurement.

Suppose 12+ repetitions for vocabulary learning to occur is sensible. In that case, the bottom row in Table 3 answers RQ2 about how many words beyond the first 3000 word families EFL learners may often encounter enough for acquisition by watching English-subtitled Mandarin dramas for one to three years during college. An average of 3999, 4285, and 4434 word families beyond the first 3000 would be met twelve times or more when they finish viewing 12, 24, and 36 Mandarin dramas respectively.

It is worth mentioning that the words appearing fewer than 12 times in a single Mandarin drama may be encountered a few more times later when watching another. When more English-subtitled Mandarin dramas are watched, the number of word families beyond the first 3000 appearing 12 times or more would gradually increase. As shown in Table 3, if EFL fans watch 12 English-subtitled Mandarin dramas, they would encounter most of the 1st to 5th 1000 word families 12 times at the minimum if we consider 'most' as at least 80% (i.e., at least 800 out of 1000 word families). Knowing 80% or more of the words at a certain 1000-word-family level indicates that one knows enough of the vocabulary at that level to focus on words from the next level (Nation, 2014). If fans continue to watch English-subtitled Mandarin dramas for another year, totaling 24 dramas, they will meet at least 800 of the 6th 1000 word families 12 times or more. When they keep viewing English-subtitled Mandarin dramas, say 36 dramas, before university graduation, they will still not meet over 800 of the 7th 1000 word families (699<800) often enough for potential learning (see Table 3 for the column of 3 years). From the 7th 1000 onwards, less than 800 words occurred 12+ times, indicating that the vocabulary recycling of English subtitles in Mandarin dramas

is vital within the first 6000 word families. Compared with graded readers for intermediate EFL learners, topping out at the 4000 word-family level, Mandarin drama English subtitles seem ideal in this regard. However, EFL learners may progress slowly beyond the 6th 1000 word-family level.

Table 3. Number of word families beyond the first 3000 word families occurring 12+ times

Viewing amount	12 Mandarin dramas	24 Mandarin dramas	36 Mandarin dramas
Attainable time frame	one year	two years	Three years
Subtitle amount (in words)	1.5 million	3 million	4.5 million
1st 1000	999	1000	1000
2nd 1000	995	1000	1000
3rd 1000	951	977	994
4th 1000	842	897	923
5th 1000	801	823	838
6th 1000	727	800	819
7th 1000	646	667	699
8th 1000	479	509	521
9th 1000	267	322	347
10th–25th 1000	237	267	287
Total	6,861	7,302	7,538
Subtotal (beyond 3000)	3,916	4,325	4,544

4.3 Learners' Perception Towards Viewing English-Subtitled Mandarin Dramas

Based on the Vocabulary Levels Test (Webb, Sasao, & Balance, 2017) scores, the recognition vocabulary of the twenty volunteers ranged from the first 3000 to 5000 word-family level. Sixteen of them converged at the 3000–4000 word-family level with three students reaching the 5000 word-family level and only one with a base vocabulary at the 2000 word-family level. Less than four months before the deadline, the twenty participants had finished viewing two English-subtitled Mandarin series of their choice.

Subsequently, they were interviewed to express their views on watching English-subtitled Mandarin series. The twenty participants showed a certain level of consistency in their responses regarding the difficulty of English subtitles. They all commented that viewing English-subtitled Mandarin dramas was not as troublesome as expected. Even the only learner with a base vocabulary at the 2000 word-family level did not find English subtitles challenging at the onset. This may be because their vocabulary size attained a minimum threshold at the first 2000 to 3000 word-family level. Through visual representations and cumulative background knowledge over episodes, knowledge of words covering less than 95% of subtitled text may still provide general comprehension of the scenarios and plots.

Unsurprisingly, most of them admitted that they did not pay full attention to English subtitles when they watched Mandarin dramas since they could hear Mandarin (L1) dialogues effortlessly. One jokingly suggested turning off the audio or reducing the volume for this issue. However, surprisingly, two participants said they turned off the audio to concentrate on reading English subtitles. They further replied that they could easily guess meanings, and the mode of thought in Mandarin dramas is predictable since Mandarin is their mother tongue.

Some other participants also seconded the two students' viewpoints about easier guessing due to familiar culture and customs. One participant pointed out the word 'waste' as an example. This participant described a dramatic scene she saw in a Mandarin drama, where a man gave a piggyback to his drunk female friend, who had just broken up with her boyfriend and did not know that the man at her side had a one-sided love for her. With his unrequited love on his back shaking violently and shouting in delirium, he strenuously climbed stairs to the rooftop hut of an old apartment without elevators, saying "Shush! You are wasted. Do you want the whole world to know you were dumped?" This participant said she already knew the usual meaning of waste, denoting use carelessly, but 'You are wasted' impressed her when she heard the Mandarin dialogue about 'You are drunk.' Thus, she learned a new meaning of 'waste' from this English subtitle. Following her remark, another student added that he was also impressed with the phrase "driving under the influence" or the abbreviation DUI, meaning drunk driving, due to frequent occurrences in Mandarin dramas. He gave his reason that he had been constantly hearing admonishments against drunk driving in daily life but had little idea how to say this offense in English. After the English subtitles showed DUI or under the influence several times, he started to become aware of this term.

Overall, the interview reveals that learners would likely uptake the words 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.

5. Conclusion

The binge-watching phenomenon among Taiwanese college students inspired this research. Because our students enjoy watching Mandarin 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 engage in watching dramas with English subtitles instead of their L1 subtitles. This preliminary study has aimed to show the potential of English subtitles as an alternative to graded readers for extensive reading.

Mandarin drama English subtitles reached the 2nd–3rd 1000 word-family levels at 95% text coverage and extended to the 4th–5th 1000 levels at 98% coverage subject to genres. This study adopted twelve repetitions as a cut-off point for incidental vocabulary learning. Results demonstrated that 3916, 4325 and 4,544 word families beyond the first 3000 word families have fulfilled this criterion if viewing 12, 24 and 36 English-subtitled Mandarin dramas respectively. This may have been a conservative assessment because learners may gain partial knowledge of the words met fewer than twelve times. Overall, EFL Mandarin drama fans can encounter most of the first 6000 word families often enough for incidental learning by watching 24 English-subtitled Mandarin dramas at the minimum. However, their vocabulary size may level off at the first 6000 word-family level. Despite this, it may be better than they do extensive reading of graded readers, topping out at the first 4000 word-family level. For advanced vocabulary growth, students can be encouraged to read English novels or news extensively.

To overcome the fear of watching English-subtitled series, learners can be advised to proceed to watch further on. After they get used to the subtitling speed, the mapping of the L2 form to L1 meaning will improve with more and more episodes being viewed, because later recurrence may consolidate the form-meaning association. There are still some other ways of helping EFL learners to reduce their vocabulary load. Gaining background knowledge through trailers and previews facilitates comprehension; therefore, learners have more leeway to concentrate on English subtitles. 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.

The goal of this research has been twofold: to measure the vocabulary level of Mandarin drama English subtitles for EFL learners to start and to estimate the potential vocabulary growth if binge-watching English-subtitled Mandarin dramas for one to three years. The results may serve as a reference for College English teachers and learners who concerned with extensive reading outside class.

Lastly, this study has focused on individual words. Despite the potential learning of the first 6000 word families, hidden in the first 6000 word families are multiword expressions, which are worth investigating but beyond the present focus. Further research may follow the present study by switching the focus from single words to multiword expressions such as semantically non-compositional phrases and idioms.

Acknowledgments

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. She is also grateful to Professor Laurence Anthony for the freeware tool AntWordProfiler.

Authors' contributions

Not applicable—sole author

Funding

Not applicable

Competing interests

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Canadian Center of Science and Education.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data are not publicly available due to copyright or ethical restrictions.

Data sharing statement

No additional data are available.

References

Anthony, L. (2023). *AntWordProfiler* (Version 2.1.0) [Computer software]. Tokyo, Japan: Waseda University. Retrieved from https://www.laurenceanthony.net/software

Bairstow, D., & Lavaur, J. M. (2011, June 15–18). *Standard vs. reversed subtitles: Effects on movie comprehension and lexical retrieval* [Paper presentation]. The 8th International Symposium on Bilingualism, Oslo, Norway.

Čepon, S. (2011). Interlingual subtitling as a mode of facilitating incidental foreign language acquisition. *English for Specific Purposes World*, *11*(33), 1–36. Retrieved from http://www.espworld.org/e.htm

Claridge, G. (2012). Graded readers: How the publishers make the grade. *Reading in a Foreign Language*, *24*(1), 106–119. Retrieved from http://hdl.handle.net/10125/66668

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, *42*(4), 497–527. https://doi.org/10.1111/j.1467-1770.1992.tb01042.x

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta* (Montréal), *49*(1), 67–77. https://doi.org/10.7202/009021ar

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge University Press. https://doi.org/10.1177/003368829802900211

Day, R. R., & Bamford, J. (2004). *Extensive reading activities for teaching language*. Cambridge University Press.

Fazilatfar, A. M., Ghorbani, S., & Samavarchi, L. (2011). The effect of standard and reversed subtitling versus no subtitling mode on L2 vocabulary learning. *Journal of Teaching Language Skills*, *30*(1), 43–64. Retrieved from https://en.jref.ir/21111

Gorjian, B. (2014). The effect of movie subtitling on incidental vocabulary learning among EFL learners. *International Journal of Asian Social Science*, *4*(9), 1013–1026. https://www.aessweb.com/journals/5007

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

Holobow, N. E., Lambert, W. E., & Sayegh, L. (1984). Pairing script and dialogue: Combinations that show promise for second or foreign language learning. *Language Learning*, *34*(4), 59–74. https://doi.org/10.1111/j.1467-1770.1984.tb00352.x

Horst, J. (2013). Context and repetition in word learning. *Frontiers in Psychology*, *4*(149), 1–11. https://doi.org/10.3389/fpsyg.2013.00149

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 1–11. https://doi.org/10.3389/fpsyg.2011.00017

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, *61*(3), 355–382. https://doi.org/10.3138/cmlr.61.3.355

Hsu, W. (2020). Can TED Talk transcripts serve as extensive reading material for mid-frequency vocabulary learning? *TEFLIN Journal*, *31*(2), 181–203. https://doi.org/10.15639/teflinjournal.v31i2/181-203

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

Kim, N. (2020). Languages on the screen: A study on reversed subtitling and EFL receptive language skills. *Multimedia-Assisted Language Learning*, *23*(1), 37–66. Retrieved from https://typeset.io/journals/multimedia-assisted-language-learning-1g5ayf2d

Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Kwarteng, A. H. (2023). Incidental vocabulary acquisition through watching movies with bimodal and standard subtitles: The case of L2 Swahili learners. *Journal for Language Teaching*, *57*(2), Article 5935.

https://doi.org/10.56285/jltVol57iss2a5935

Lambert, W. E., Boehler, I., & Sidoti, N. (1981). Choosing the languages of subtitles and spoken dialogues for media presentations: Implications for second language education. *Applied Psycholinguistics*, *2*(2), 133–148. https://doi.org/10.1017/S0142716400000904

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? *The Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16. Retrieved from http://hdl.handle.net/10125/66881

Nation, I. S. P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, *27*(1), 136–145. Retrieved from http://hdl.handle.net/10125/66705

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins. https://doi.org/10.1075/z.208

Nation, I. S. P. (2017). *The BNC/COCA Level 6 word family lists* (Version 1.0.0) [Data set]. Retrieved from 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 ed., pp. 34–52). Edward Arnold.

Netflix. (2014). *Netflix declares binge-watching is the new normal*. Retrieved from 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

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

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 University Press. https://doi.org/10.1017/CBO9780511732942

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

Subtitles. (2024). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Subtitles

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

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. Retrieved from 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

Zanon, N. T. (2006). Using subtitles to enhance foreign language learning. *Porta Linguarum*, *6*, 41–52. https://doi.org/10.30827/Digibug.30659

Zanon, N. T. (2007). Learning vocabulary through authentic video and subtitles. *Tesol-Spain Newsletter*, *31*, 5–8. Retrieved from https://www.tesol-spain.org/en/