



## Korean Drama Fever—Expanding English Lexicon through Watching English-Subtitled K-Dramas: The Case of Non-Compositional Multi-word Expressions

**Wenhua Hsu**

whh@isu.edu.tw

I-Shou University, Kaohsiung City, Taiwan

Received: 25 March 2023

Accepted: 9 May 2023

Published: 20 June 2023

DOI: <https://doi.org/10.33541/jet.v9i2.4761>

### Abstract

This research was prompted by the phenomenon of binge-watching Korean television series (K-drama) amongst college students in Taiwan, where English as a foreign language (EFL) is a required course. The researcher-teacher sought to create a pedagogically useful list of the frequent semantically non-compositional multi-word expressions (MWEs) for EFL learners with K-drama fever who often binge-watch K-dramas. A corpus of 25+ million English subtitled words derived from 240 K-dramas across different genres was compiled. Based upon a set of criteria (frequency, range, meaningfulness, well-formedness, non-decomposability and semantic non-compositionality), a total of 326 MWEs of 2 to 6 words were selected. The 326 phrasal expressions are mostly composed of the first 3000 word families. As with other individual word lists, it is hoped that the listing of the non-compositional MWEs may serve as a reference for General English teachers.

### Keywords:

*multi-word expressions; semantic non-compositionality; English subtitles; K-dramas*

### INTRODUCTION

Korean Wave (coined by Beijing journalists to describe the surging popularity of Korean entertainment) started in the mid-1990s and has had a huge base of global followers for three decades (Jin & Yoon, 2017; Lee & Nornes, 2015). This study was prompted by the phenomenon of binge-watching Korean television series (K-drama) amongst college students in Taiwan, where General English is a required subject for freshmen and sophomores. The researcher-teacher often overhears her students sharing information about which K-drama they have been binge-watching recently. The binge-watching phenomenon began in the late 2000s thanks to the rapid rise of streaming platforms, which can be accessed anytime, anywhere worldwide. In a market survey by Netflix (2013),

about 1500 TV streaming viewers depicted their binge-viewing behavior as watching 2 to 6 episodes of the same program in one sitting.

In view of K-drama fever, the researcher-teacher is concerned with English subtitles as learning input, given that our students can change their viewing habit from Chinese (the first language, L1) to English subtitles. As per Markham and Peter (2003), L2 video subtitles are on-screen text in the viewer's L1. Different from captioned videos (L1 videos with L1 subtitles), the current context refers to subtitles in the viewer's L2 (English) with L3 (Korean) video, which may compel non-Korean viewers to depend on screen text heavily. On over-the-top (OTT) streaming platforms (e.g. Rakuten Viki), K-dramas are foremost dubbed into English, and then English subtitles are translated into many other languages (Locher & Messerli, 2020; Pedersen, 2019).

Subsequent to THE AUTHOR's (in press) research regarding the vocabulary levels that K-drama English subtitles involve and the vocabulary learning opportunities they afford, this research switched the focus from single words to multi-word expressions. According to THE AUTHOR (in press), K-drama English subtitles reached the 2000—3500 word-family levels at 95% text coverage and extended to the 4000—5500 levels at 98% coverage subject to genres. As with extensive reading of graded readers, EFL K-drama fans can encounter most of the first 5000 word families often enough for learning to occur through continually watching English-subtitled K-dramas during college years.

Despite the potential learning of the most frequent 5000 word families, concealed in the first 5000 word families are multi-word expressions (hereafter MWEs). Some MWEs may go unnoticed or misinterpreted, especially when learners presume that they have acquired mastery of high-frequency words (e.g. *piece, fit, have, a, the, out, of, work, blue*) but actually they are not familiar with their multi-word combinations (e.g. *have a fit; a piece of work; out of the blue*). As shown, these MWEs may cause deceptive comprehension (Martinez & Murphy, 2011), if not known. As such, particular attention is paid to semantically non-compositional MWEs as the researcher reasoned that they form semantic units to express specific concepts and can be learned like single words. This research addressed the following two questions in a bid to expand K-drama buffs' English lexicon in an EFL setting.

1. By watching English-subtitled K-dramas, what are the semantically non-compositional MWEs that EFL K-drama fans may encounter often enough for potential learning to occur?
2. What discourse functions do non-compositional MWEs perform in dramas lines?

## LITERATURE REVIEW

### Multiword Expressions (MWEs)

Masini (2005) defined MWEs as “lexical units larger than a word that can bear both idiomatic and compositional meanings” (p. 145), while Hinkel (2023) referred to them as “recurrent combinations of words — words that are connected to other words — that are remembered and used as single lexical [vocabulary] items (p.2).

MWEs are ubiquitous and make up a large proportion of any discourse (Nattinger & DeCarrico, 1992). Native speakers may have thousands of “lexicalized sentence stems” at their disposal (Pawley & Syder, 1983, p. 214). Altenberg (1998) gauged that various MWEs account for as high as 80% of the words in the London-Lund Corpus while Erman and Warren (2000) estimated that prefabricated MWEs make up 55% or more of the

words in a text. Individual words are merely the tips of phraseological icebergs (Martinez & Schmitt, 2012).

Some words frequently co-occur with other words and form relatively fixed multi-word sequences. This phenomenon is generally referred to as formulaic language, and each individual instance of formulaic language is called a formulaic sequence (Schmitt, 2010). Wray (2002, p. 9) defined a formulaic sequence as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.” This definition indicates that formulaic sequences behave much like individual words, stored in the mental lexicon and used as a non-decomposable unit.

Due to lack of a universal definition for recurrent MWEs, they have been labeled in a range of ways: *collocations* (Altenberg, 1998; Howarth, 1998), *lexical bundles* (Biber, Conrad, & Cortes, 2004; Hyland, 2008), *clusters* (Scott, 1996), *formulaic sequences/formulae* (Martinez & Schmitt, 2012; Wray, 2002), *sentence stems* (Pawley & Syder, 1983), *prefabricated units/prefabs* (Cowie, 1998), *n-grams* (Stubbs, 2007) and frozen phrases (Wood, 2020). Similarly, idioms, phrasal verbs, proverbs, and binomial expressions display one facet of formulaic language respectively.

On account of the multiplicity of formulaic language, the defining characteristics vary from researcher to researcher. This study used multi-word expressions (MWEs) as an umbrella term to refer to miscellaneous combinations of words, involving different degrees of semantic compositionality and syntactic fixedness. If the interpretation of a MWE can be derived from the meanings of its component words, it is semantically compositional. Conversely, it is a non-compositional MWE if its individual words do not help each other to reveal its meaning as a whole.

### **Discourse Functions of MWEs**

Biber, Conrad, and Cortes (2004) designed a categorization scheme for lexical bundles commonly used in university spoken and written registers. There are four categories according to discourse functions: referential, discourse-organizing, stance and interactional bundles. Referential bundles perform an ideational function and the signals they send involve location, time, quantity and procedure as well as the description of attributes. Discourse-organizing bundles are concerned with transition signals to show relationships among ideas. Stance bundles express attitudes, assessments or propositions, while interactional bundles are used to engage listeners in participation.

Based on 238 idiom types gleaned from the MICASE, Simpson and Mendis (2003) enumerated six discourse functions. They are used for evaluation, description, paraphrase, emphasis and collaboration for shared views as well as used in metalanguage such as discourse organizing.

In this research, the above taxonomies are pedagogically helpful in raising learners' awareness of the discourse functions that semantically non-compositional MWEs perform in drama lines.

### **Lexical Text Coverage and Number of Repetitions for Lexical Learning**

Nation (2006) defined lexical text coverage as “the percentage of running words in the text known by the reader” (p. 61) and advocated that 98% coverage (2 unknown words per 100 words) is ideal for guessing words from context and may provide good conditions

for lexical learning. The putative 98% lexical text coverage has been widely adopted as a benchmark for adequate comprehension (Nation, 2006), for pleasure reading (Hirsh & Nation, 1992) as well as for independent/unassisted reading (Hu & Nation, 2000).

When lexical text coverage with an emphasis on individual words is calculated, MWEs are not taken into account. As a result, the lexical coverage of a text may be overestimated when semantically non-compositional MWEs are hidden in known words and their meanings as a whole happen to be unfamiliar to learners. Accordingly, knowledge of non-compositional MWEs would contribute to filling the rift with text coverage that individual words fail to account for (Martinez & Murphy, 2011).

Previous research on word learning has documented that a single encounter with a new word seldom supports robust learning of it (Horst, 2013; Horst, Parsons & Bryan, 2011). Nation (2014) inferred from past studies that it takes from 5 to 16 exposures to a word or a chunk for uptake to happen, and assumed that 12 encounters with a word in a variety of contexts would just be enough to develop knowledge of that word. Following Nation (2014), the researcher adopted 12 times as a cutoff frequency to cull non-compositional MWEs for inclusion in the list.

### Methods to Identify MWEs

It is generally agreed that frequency is a good indicator in deciding the usefulness of a lexical item in terms of learning returns. The pre-determined cutoff values for frequency have been arbitrary, depending upon researchers' discretion. Biber, Johansson, Leech, Conrad and Finegan (1999) adopted a cutoff point at occurring at 10+ times per million words. Cortes (2004) opted for 20 times, when comparing the functions of lexical bundles used in history and biology writings. Biber, Conrad and Cortes (2004) were more rigorous in selecting lexical bundles by setting a relatively high cutoff at 40 times per million tokens.

Resorting to frequency alone, n-gram extraction tools may generate thousands of multi-word combinations, some of which are not "pedagogically compelling" (Simpson-Vlach & Ellis, 2010, p. 493). For instance, *'the one who took'* and *'else could it be'* are 4-word bundles without complete meanings. Straddling two phrasal boundaries, some lexical bundles are not readily accessible for teaching and learning (e.g. *some kind of a, how could something like the*).

Apart from frequency-based retrievals, Simpson-Vlach and Ellis (2010) considered another quantitative measure and proposed the idea of Formula Teaching Worth (FTW). In their endeavor to compile an Academic Formulas List, both frequency and mutual information (MI) were factored in multiple regression analyses. MI is a statistical measure for cohesiveness of words, indicating collocation strength (Stubbs, 1995). Multiword combinations with high MI values are more likely to be meaningful and are therefore worth pedagogical attention. Simpson-Vlach and Ellis (2010) concluded that the FTW metric that combines frequency and MI may provide teachers with prioritizing criteria, when judging multi-wordsequences in terms of whether they are worthy of instruction.

To identify the most frequent spoken collocations for deliberate learning, Shin and Nation (2008) used six criteria and underwent laborious manual inspection. Among a series of criteria that they applied was "grammatical well-formedness" (p. 341). They targeted collocations which do not span two "immediate constituents" (Bloomfield, 1933, p. 161), since a well-formed MWE is a comprehensible unit. For example, *'extent that the'* is less understandable than *'to the extent that'*.

In consideration of semantic non-compositionality, Martinez and Schmitt (2012) attempted to identify the most frequent opaque formulaic sequences (OFSs) that can be learned as lexical units and can be integrated into the 1<sup>st</sup> to 5<sup>th</sup> 1000 word-family lists along the British National Corpus (BNC) word-frequency scale. They established six criteria to minimize intuitions in deciding whether a multi-word expression is a Morpheme Equivalent Unit and potentially “deceptively transparent” (Laufer, 1989, p. 11).

Through statistical computation and the judgement of a panel of experts for refining selection, Ackermann and Chen (2013) retrieved 2,468 most frequent lexical collocations from the 25-million-word Pearson International Corpus of Academic English to help students increase their academic collocation competence. A look at the ACL shows that a great many semantically compositional collocations (e.g. *academic writing*, *online database*, *further research*) are already within our students’ grip and may not be their imminent concern. In view of the fact that not all MWEs are of equal importance to learners, this research adopted semantic non-compositionality as a point of departure.

## METHOD

### The Corpora

The present corpus contained the K-dramas with high viewership ratings according to OTT media services and Nielsen Korea. For example, *The World of the Married*, *Reborn Rich* and *Sky Castle* were the three highest-rated miniseries on cable TV (JTBC, tvN, ENA, OCN, etc.) as of 2022. This implies that these dramas are likely to have been viewed by numerous K-drama fans. Public channels’ (KBS, SBS, MBC) high-rated dramas were also included. For instance, *Jewel in the Palace* with 60.8% nationwide viewership has been broadcasting in 150 countries so far.

Moreover, drama genres were taken into account in order to build a comprehensive corpus of K-dramas (see Table 1). K-drama storylines often present some social issues that audiences from all over the world can relate to, such as wealth inequality, sexual harassment, bullying and corruption as well as a wide range of topics from all walks of life. The diverse subject matters suggest that K-dramas, given English subtitles, may be a rich resource for real-life language learning.

K-drama English subtitles in SRT format were downloaded from the Internet for research purposes (as below).

9  
00:02:28,064 --> 00:02:29,399  
It was the last tile.

10  
00:02:31,067 --> 00:02:32,610  
He was clearly going to move.

11  
00:02:32,694 --> 00:02:33,903  
How can you be so sure?

12

00:02:35,071 --> 00:02:37,907

He knew how to tell apart  
the tempered glass,

13

00:02:37,991 --> 00:02:40,910

but just watched people fall down and die.

14

00:02:40,994 --> 00:02:44,747

In the end, both you and I  
crossed that bridge alive thanks to him.  
(from *Squid Game*, Episode 8)

It should be noted that the English subtitles of a K-drama provided by different OTT streaming services may not be exactly the same or may not have total accuracy [see Hall (2021) for a comparison of translation by paid translators for Netflix and by volunteer fansubbers for Rakuten Viki]. Even so, the present corpus, containing 25+ million subtitled words from 240 K-dramas, may be large enough to provide reliable assessments.

Ten sub-corpora (by genre) with 12 K-dramas in each were built for reliability check. The inclusion of 12 K-dramas in each sub-corpus was based upon the assumption that it is feasible for K-drama fans to watch one drama with 16 to 24 episodes per month, 12 K-dramas per year. However, viewers may not stick to the same genre all the time. A lot of audiences watch K-dramas based on netizens' recommendations, while a great many fans follow their favorite stars (e.g. Hyun Bin *현빈*, Song Joong-ki *송중기*) regardless of genres. In consideration of these two binge-watching behaviors, another two genre-mixed sub-corpora were added (see Table 1).

### Procedures

*COLLOCATE* (Barlow, 2004) was used to retrieve MWEs from the 25-million-token K-drama English subtitle corpus. The span parameter for word length was set from 2 to 6, because frequencies drop drastically as MWEs are extended to five words or beyond (Hyland, 2008).

The cutoff frequencies for the selection of lexical bundles in past studies ranged between 10 and 40 times per million tokens. To prevent important MWEs from being excluded at the initial stage, 10 times per million tokens in this research was set to begin with, namely 12 times per 1.2 million tokens (roughly equivalent to 12 K-dramas, see Table 1 for the total tokens of each sub-corpus).

As aforementioned, learning rarely occurs after a single encounter. As the number of encounters with an unknown word or chunk increases, the potential of learning that word or chunk increases. Following Nation (2014), the researcher chose 12 times as a threshold and measured the number of non-compositional MWEs appearing 12+ times in English subtitles.

Table 1

*Sizes of the K-drama Corpora*

Corpus	Number of K-dramas	Total tokens
The main corpus	240	25,967,735
Sub-corpora	12 K-dramas in each sub-corpus	

Coming of age	<i>Twenty-five, twenty-one, All of Us are Dead, Fight for My Way, the Heirs, Dream High, etc.</i>	1,402,357
Historical	<i>Under the Queen's Umbrella, The Red Sleeve, Moon Embracing the Sun, Jewel in the Palace, etc.</i>	1,398,896
Time travel	<i>Mr. Queen, Go Back Couple, Signal, Nine Time Travels, The King: Eternal Monarch, etc.</i>	1,323,344
Thrillers	<i>Flower of Evil, Kingdom, Penthouse, Beyond Evil, Hell is Other People, Sweet Home, Hellbound, etc.</i>	1,389,778
Crimes	<i>My name, Voice, Stranger, Mouse, Tunnel, Awaken, Watcher, etc.</i>	1,399,873
Medical	<i>Hospital Playlist, Doctor Romantic, Doctor Prisoner, Doctor John, Good Doctor, etc.</i>	1,357,213
Legal	<i>One Dollar Lawyer, Innocent Defendant, Hyena, etc.</i>	1,466,432
Fantasy	<i>Alchemy of Souls, Guardian-The Lonely and Great God, My Love from the Star, Hotel del Luna, W:Two Worlds, etc.</i>	1,395,774
Action	<i>Again My Life, Vincenzo, Vagabond, City Hunter, IRIS, etc.</i>	1,289,347
Romance	<i>Business Proposal, It's Okay to Not Be Okay, What's Wrong with Secretary Kim, Her Private Life, True Beauty, etc.</i>	1,333,276
Popular stars	<i>Reborn Rich, Descendants of the Sun, Crash Landing on You, Secret Garden, Itaewon Class, Big Mouth, While You Were Sleeping, etc.</i>	1,377,456
Netizens' recommendations	<i>Extraordinary Attorney Woo, Dazzling, My Mister, Reply 1988, Taxi Driver, Move to Heaven, The World of the Married, Our Blues, Prison Playbook, Sky Castle, etc.</i>	1,376,499

Since one of the goals was to identify the non-compositional MWEs that commonly appear in English subtitles, those that occurred with a very high frequency but in only one or two drama genres would not be taken into account. Specifically, MWEs in different inflectional forms taken together had to appear in each of the twelve sub-corpora across different genres. The decision was admittedly arbitrary but in line with the present goal of widespread use.

Another consideration for selection was meaningfulness. The recurrent MWEs must have meanings and can be learned as single units. This principle would make them comparable to a list of individual words. Prior to manual vetting, *Mutual Information* (MI) was used to filter out free word combinations. According to Hunston (2002), multiword combinations with the MI value greater than 3 are considered strong in terms of cohesiveness. MWEs with the MI value less than 3 were deleted at this stage. Examples include, *buthow, who's in*.

Referring to Martinez and Schmitt (2012) as well as Shin and Nation (2008), the researcher formulated seven questions to guide the decision of candidate MWEs for inclusion in the list. They were used to determine meaningfulness (Q1), well-formedness (Q2), non-decomposability (Q3) and semantic non-compositionality (Q4 to Q7).

Q1. Does the candidate MWE convey a meaning?

Q2. Does the candidate MWE span two phrasal boundaries?

Q3. Does the candidate MWE behave like an individual lexical item, which is unlikely to be further decomposed into its subparts?

Q4. Does the meaning of the candidate MWE still remain when each component word is decoded with its core meaning?

Q5. Does the candidate MWE have more than one meaning?

Q6. Does the candidate MWE contain a word with more than one meaning?

Q7. Does the candidate MWE contain a word with a derivational affix which alters the meaning of its base form?

From Q1 to Q7, the researcher and her colleague separately made a judgment on approximately 58,950 candidate MWEs with 12+ repetitions and  $MI > 3$ . The responses of *yes*, *not sure* and *no* were coded as 1, 0.5 and 0 respectively to form a 3-point scale. When there was no consensus, the entry was decided for tentative exclusion from the candidate MWEs list, subject to further confirmation. Then a series of Cohen's Kappa statistics for each question were undertaken to check inter-rater reliability. The  $k$  values were all greater than 0.8, showing a substantial agreement between the two raters.

To judge whether the MWEs were meaningful, well-formed and non-decomposable, they were reviewed against Q1, Q2 and Q3 respectively. For semantic non-compositionality, Qs 4 to 7 were used to examine the remaining MWEs. Polysemous MWEs or those with one word having multiple meanings may be problematic for learners. The cases in point of the former are *worked up*, *a piece of work*, *fall for*, *up to* and *know a thing or two*, while the instances for the latter are *fit in* *throw a fit*, *sorts in* *out of sorts*, *wasted* (drunk) in *get wasted*, *break in* *give me a break* and *foot in* *my foot* (my gosh). Q5 and Q6 were used to pursue the MWEs with multiple meanings either in one constituent word or as a whole.

It was found that some candidate MWEs with a word containing a derivational affix may mislead learners into making a wrong form-meaning link. Learners may think that they know the base form of the word, but they are unaware that the meaning of its derivational form has been altered (e.g., *could hardly*, *profession of love*, *I am screwed*). Q7 was therefore supplemented for screening.

### Data Processing

In the compilation of non-compositional MWEs, a few modifications were made. One was modified for MWEs in different inflectional forms (e.g., -s, -ing, -ed). They were combined to form a single item with their lemma as the representative form. An example is *give somebody an earful* (a sum of 46 times) = *gives \* an earful* (occurring 1 times) + *gave \* an earful* (24) + *given \* an earful* (3) + *giving \* an earful* (12) + *give \* an earful* (6). This phrasal expression was entered in the search bar as *[give] \* an earful*, with *[give]* instructing the search engine to look for the lemmas of the verb *give* (gives, gave, given, giving and give). 'Somebody' was replaced by the wildcard character \* denoting any token between *give* and *an earful*. The assumption was that focusing on a single entry at a time (*give somebody an earful* in this case) may be less complicated for students to learn at the onset. After they have some familiarity with its salient meaning (to give one a lengthy reprimand or lecture), its variants (*gives/gave/giving you/him/her/them/us an earful*) may be acquired with more exposure later.

Another modification was adding the copula *be* to verb past participle + preposition combinations if they can be used as predicate, for example, *be worked up* (=excited and angry), *be supposed to* and *be taken aback*. The other revision was made for partial overlap, where a shorter MWE was subsumed in a longer one, for example, *as well* in *as well as*. To obtain an accurate frequency of *as well*, subtractions were made from the frequency of *as well as* ( $3832 - 390 = 3442$ ). Since these two phrases can stand alone as a meaningful unit, they were separately compiled into the MWE list.



To examine multiple meanings, a polysemous MWE was checked with the Free Dictionary (<http://idioms.thefreedictionary.com>). A statistical tool can calculate the total number of occurrences of the same form but cannot detect its different meanings from context and figure out the frequency of each meaning. For instance, *to death* occurred 56 times in the present corpus, including the occurrences of the meanings *extremely* and *until dead*. Its total frequency was tallied instead of the individual frequency of each meaning.

## FINDINGS AND DISCUSSION

### Frequent Non-Compositional MWEs in K-drama English Subtitles

A total of 326 non-compositional MWEs of 2 to 6 words were ultimately selected (see Appendix). There are 163 two-word, 107 three-word, 40 four-word, 12 five-word and 4 six-word MWEs, which EFL K-drama fans may encounter often enough for potential learning to occur, if they continually watch English-subtitled K-dramas regardless of genres.

Installed with the ranked twenty-five 1000-word-family lists derived from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) (Nation, 2017), *RANGE* (Heatley, Nation & Coxhead, n.d.) was used to examine the vocabulary levels of the 326 MWEs (see Table 2).

Table 2

*Vocabulary Levels of the 326 Non-Compositional MWEs*

BNC/COCA vocabulary levels	Number of tokens	% coverage in tokens	Number of word families
1 <sup>st</sup> 1,000	857	90.12%	267
2 <sup>nd</sup> 1,000	48	5.05%	38
3 <sup>rd</sup> 1,000	11	1.16%	11
4 <sup>th</sup> 1,000	4	0.42%	4
5 <sup>th</sup> 1,000	5	0.53%	5
6 <sup>th</sup> 1,000	7	0.74%	6
7 <sup>th</sup> 1,000	1	0.11%	1
8 <sup>th</sup> 1,000	2	0.21%	2
9 <sup>th</sup> 1,000	2	0.21%	2
10 <sup>th</sup> 1,000	0	0.00%	0
11 <sup>th</sup> –25 <sup>th</sup> 1,000	14	1.45%	12
Total	951	100%	348

The 326 non-compositional MWEs comprise 951 words and involve 348 word families. The BNC/COCA first 1000 word families account for 90.12% of the total words and the second 1000 make up 5.05%, followed by the third 1000 being 1.16% coverage. The combined coverage of the first 3000 word families is 96.33%. After the first 3000 word families, the coverage of additional 1000 word families rapidly reduces to less than 1%. That is, the 326 non-compositional MWEs are composed of very general words (e.g. *seeing each other*, *I got you*, *what's with you*). Along with their high-frequency component words, the 326 MWEs occur across a wide range of drama genres. Examples include *nothing to do with*, *upper hand*, *pull yourself together*, *a big deal*, to name but a few.

The strings of content words and function words form a common pattern in the non-compositional MWE list, for example, *in order to*, *there is/are* and *even if*. As the instances have shown, the everyday words are constituent parts of a repertoire of multiword combinations that make up a discourse, as Sinclair (1991) has claimed.

Concerning the structure of 2-word MWEs, more than a half of them (91 out of 163) are grammatically-conditioned pairs, namely a content word combined with a function word (e.g. *could hardly*, *from scratch*), as opposed to 72 lexical collocations (e.g. *so dense=stupid*, *cold shoulder*, *no matter*, *gold digger*, *buy time*, *loan shark*). Amid grammatical collocations, phrasal verbs are in the majority (34/91) (e.g. *mess with*, *hit on*, *run into*, *account for*, *deal with*, *give up*, *pass out*, *black out*) and phrasal prepositions come second (18/91) (e.g. *as to*, *apart from*, *according to*, *along with*), followed by a preposition plus a noun (12/243=8.2%) (e.g. *behind bars*, *at once*, *at times*).

One pattern of the 3-word MWEs is a passive verb followed by a preposition requiring a noun phrase or by an infinitive-to for completion. For completeness sake, they are presented as *be* + past participle + preposition, as in the cases of *be bound to*, *be ripped off*, *be dumped* and *be taken aback*. When the verb-*be* is added, they form the passive and can stand alone appearing in an independent clause. The three patterns as ~ *as*, a ~ *of*, and *by* + noun phrase are also productive among the 3-word MWEs, as in the cases of *as long as*, *as soon as*, *a couple of*, *by any chance*, *by means of*, and *by way of*. These three patterns contribute to the description of quantity or an approach.

As to 4-word MWEs, the prepositional phrase is a common structure. They are, for instance, *of one's own accord*, *after one's own heart*, *out of your mind*, *out of the blue*. The non-compositional MWE list also contains quite a few idioms (e.g. *upper hand*, *put my life on the line*, *play hard to get*). Along with their high-frequency component words, these idioms occur across a wide range of topic areas in daily conversation. More examples include *a big shot*, *cross paths*, *death wish*, *gold digger*, *table death*, *dig up dirt on*, *been there*; *done that*, to name but a few.

The present MWE list is similar to Martinez's (2012) 505 phrasal expression list in terms of semantic non-compositionality. However, only 63 out of the 326 MWEs overlap with 505 phrasal expressions. This may be due to different sources of data. The present corpus was restricted to drama lines, whereas Martinez's was derived from the BNC, which contains 90% written language and 10% spoken language. Beyond single words, the 326 frequently-occurring MWEs may be another cohort of lexical items for learning for EFL learners with K-drama fever, given constant exposure to English-subtitled K-dramas.

### Discourse Functions of the Non-Compositional MWEs

The 326 non-compositional MWEs are multifaceted and it may not be easy to fold them into a compact categorization. Referring to Biber, Conrad and Cortes (2004) as well as Simpson and Mendis (2003), the researcher generalized the 326 non-compositional MWEs into five types: (1) referential uses, including informing and interpreting purposes, (2) evaluative uses, including expressing attitudes, such as debates of stances and comments, (3) emphasizing/highlighting uses (e.g. presenting a contrast of opinions), (4) discourse-organizing uses for coherence, and (5) idioms in association with imagery or vividness (see Table 3).

Table 3

#### *Distribution of the 326 Non-Compositional MWEs across Five Discourse Functions*

Hsu: Korean Drama Fever—Expanding English Lexicon through Watching English-Subtitled K-Dramas: The Case of Non-Compositional Multi-word Expressions

Discourse functions	Number of MWEs	Examples
Referential	92	according to; in terms of; in accordance with; in order to; all sorts of; a couple of; nothing to do with
Evaluative	38	be likely to; make no sense; drive me crazy; dead and buried; play hard to get; gold digger; are you looking down on me? what's with
Emphasizing	103	as soon as; lose one's mind; break up with; on top of that; as long as; by any means; come clean; of one's own accord; won't let it slide
Discourse-organizing	11	so that; in order to; as well as; by any chance
Idioms	82	loan sharks; a piece of cake; drive under the influence; cross paths; the wind is blowing; flesh and blood; work one's ass off

Among the 326 non-compositional MWEs, there are about 92 referential MWEs, 38 MWEs used for evaluative purposes, 103 for emphasis, 82 for vivid image or irony and 11 discourse-organizing MWEs. It may be challenged on the precision of categorization because a single MWE may perform more than one discourse function and a clear-cut distinction of them may be fruitless. This preliminary typology was used to explore a general pattern concerning the usage of non-compositional MWEs in drama lines.

## CONCLUSION

The principal concern of this study was to create a pedagogically useful list of frequent non-compositional MWEs for EFL learners with K-drama fever who constantly binge-watch K-dramas. By means of a principled set of criteria, a total of 326 non-compositional MWEs of 2 to 6 words were selected. The frequent MWEs are mostly made up of the first 3000 word families. Therefore, they can partially bridge the gap between the text coverage that general words can and cannot account for in English subtitles to facilitate viewing comprehension. The non-compositional MWE list is short but may be a viable option for EFL students to learn within a short period of time.

This research has identified the discourse functions that the non-compositional MWEs perform in drama lines and they may help in raising learners' awareness of how they behave in authentic discourse. Despite arbitrary decisions on cut-off values in the compilation, there may be some advantages to overt instruction of these phrases, which are worth investigation but beyond the present focus. As with other individual word lists, it is hoped that the listing of the frequent non-compositional MWEs may serve as a reference for General English teaching and learning.

## REFERENCES

Hsu: *Korean Drama Fever—Expanding English Lexicon through Watching English-Subtitled K-Dramas: The Case of Non-Compositional Multi-word Expressions*

- Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL)-A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235-247.
- Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford: Oxford University Press.
- Barlow, M. (2004). Collocate [Computer software]. Houston: Athelstan. Available from [http://athel.com/product\\_info.php?products\\_id=29](http://athel.com/product_info.php?products_id=29)
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson.
- Bloomfield, L. (1933). *Language*. New York: Henry Holt.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.
- Cowie, A. (1998). (Ed.). *Phraseology: Theory, analysis, and applications*. Oxford: Oxford University Press.
- Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text, 20*(1), 29-62.
- Hall, J. (2021). Professional vs non-professional translation of Korean TV dramas. *Neke The New Zealand Journal of Translation Studies, 4*(1), 1-42.
- Heatley, A., Nation, I. S. P., & Coxhead, A. (n.d.). RANGE [Computer software]. Retrieved from <http://www.victoria.ac.nz/lal/about/staff/paul-nation>
- Hirsh, D., & Nation, I.S.P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689-696.
- Horst, J. (2013). Context and repetition in word learning. *Frontiers in Psychology, 4*(149), 1-11.
- 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.
- Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics, 19*(1), 24-44.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.
- Jin, D., & Yoon, T. (2017). The Korean wave: Retrospect and prospect—Introduction. *International Journal of Communication, 11*, 2241-2249.
- Hinkel, E. (2023). Teaching and learning multiword expressions. In E. Hinkel (Ed.), *Handbook of practical second language teaching and learning* (pp. 435-448). Routledge.
- 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). Clevedon, England: Multilingual Matters.
- Lee, S., & Nornes, M. (2015). (Eds.). *Hallyu 2.0: The Korean wave in the age of social media*. Ann Arbor: University of Michigan Press.

- Locher, M. A., & Messerli, T. C. (2020). Translating the other: Communal TV watching of Korean TV drama. *Journal of Pragmatics*, 170(1), 20-36.
- 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.
- Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, 45(2), 267-290.
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320.
- Masini, F. (2005). Multi-word expressions between syntax and the lexicon: The case of Italian verb-particle constructions. *SKY Journal of Linguistics*, 18, 145-173.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82. 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.
- Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- Nattinger, J. R., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Netflix. (2013). *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>
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-225). London: Longman.
- Pedersen, J. (2019). Fansubbing in subtitling land: An investigation into the nature of fansubs in Sweden. *Target*, 31(1), 50-76.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Hampshire, England: Palgrave Macmillan.
- Scott, M. (1996). WordSmith tools [Computer software]. Oxford: Oxford University Press. Available from <https://lexically.net/wordsmith/downloads/>
- Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62(4), 339-348.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37(3), 419-442.
- Simpson-Valch, R., & Ellis, N.C. (2010). An academic formulas list: New methods in phraseology Research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stubbs, M. (1995). Collocations and semantic profiles: on the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 23-55.
- Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 years on* (pp. 89-105). Amsterdam: Radopi.
- Wood, D. (2020). Classifying and identifying formulaic language. In S. Webb (Ed.), *Routledge handbook of vocabulary studies* (pp. 30-45). Oxon: Routledge.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

**APPENDIX**

**326 Frequent Semantically Non-Compositional MWEs Derived from K-drama English Subtitles**

<i>2-word</i>	<i>3-word</i>
a bit	(be) bound to
a blackout	(be) concerned about
a few	(be) entitled to
a hotshot	(be) faced with
a little	(be) likely to
a lot	(be) ripped off
above all	(be) supposed to
according to	(be) taken aback
account for	(be) worked up (excited, angry)
all along	(I) got you
all over	(make) no mistakes
along with	a big deal
apart from	a big shot
as if	a bit much
as to	a couple of
as usual	a crush on
as well	a dead meat
at all	a good/ great deal of
at least	a handful of
at once	a little bit
at times	a lot of
back then	a number of
bar exam	a rainy day
behind bars	a variety of
better off	all kinds of
black belt	all sorts of
black box	as far as
black out	as long as
blind date	as much as
bring up/ be brought up	as soon as
buy time	as well as
carry on	be about to
carry out	break up with
come clean	by any chance
come on	by any means
come true	come up to
could hardly	come up with
cross paths	cut the crap
deal with	dead and buried
death wish	drive me crazy/nuts
dig around	flesh and blood
drop dead!	for goodness' sake/ for god's sake
due to	get away with
each other	get on with
even though	get out (of)

---

ever since	get rid of
excuse me	go easy on
fall for	grab a bite
fall through	grab a drink
feel like	hang out with
figure out	happen to (be)
fond of	have/ throw a fit
forget it	have feelings for
freak out	have got to
from scratch	head over heels
game over	hit on (girls/women)
get along	I'm afraid
get changed	I'm doomed.
get dumped	I'm screwed.
get going	in a row
get lost	in accordance with
get wasted (=drunk)	in addition (to)
give up	in broad daylight
go ahead	in charge (of)
go on	in order to
go overboard	in return (for)
gold digger	in terms of
green light	in touch (with)
had better/ 'd better	let go (of)
hang on	let it slide
have to	Let's grab
hold onto	let's see
if only	look forward to
in case	lose one's mind
in common	lower/let down one's guard
in time	make ends meet
instead of	none other than
kind of	not work out
live with	nothing more than
loan shark(s)	on behalf of
look after	on the sidelines
look for	one more time
look into	out of concern
lots of	out of order
love shot	out of the blue
lovey-dovey	pack one's bags
make sense	play around with
mess with	point of view
my goodness/ my foot/ my lord/ my gosh	pull yourself together
never mind	put up with
next door	quite a character
next to	run out of
no doubt	over and over
no idea	seeing each other
no longer	should've/ should have been
no matter	slip of the tongue
no object	some kind of

---

no sense	take care of
no way	take for granted
no wonder	take into account
noblesse oblige	take part in
of course	the rest of
off guard	the sight of
on board	the third party
on earth	what's with
on purpose	what's wrong
once again	<hr/> <b>4-word</b> <hr/>
once more	(drive) under the influence
other than	a favor to ask
out of	a piece of cake
pass away	a piece of trash
pass out	a piece of work!
pick up	after one's own heart
play dumb	all over the place
point fingers	beat around the bush
prior to	been there, done that
rather than	dig up dirt on
reek of	every nook and cranny
rely on	for the sake of/ for one's sake
rip off	give it a try/ shot
run into	give sb. an earful
shut up	give someone a break
sick of	have a hard time
slack off	in the same boat
slush fund	it seems as though
snail bride	lay a hand on
so dense (=stupid, ignorant)	make up one's mind
so far	no choice but to
so that	none of one's business
sort of	not make any sense
sort out	nothing to do with
stand for	of one's own accord
stand surety	on the same page
straight away	on top of that
straight face	once upon a time
table death	out of one's mind
take off	play hard to get
take on	pour/throw cold water on
take over	putting on an act
take place	take it out on
tend to	that sort of thing
thanks to	the wind is blowing
the breadwinner	turn a blind eye
the former	What brings you here?
the latter	what's going on
the odds	won't let it slide
there is/ there are	work one's ass off
third party	<hr/> <b>5-word</b> <hr/>
to blame	a lot on one's mind



---

to date	a stab in the back
to death	Are you playing with (me)?
to do with	get out of the/ one's way
turn down	have been through a lot
turn out	it doesn't matter to me
ulterior motive	know a thing or two
up to	no matter what it takes
upper hand	put sb./something on the line
used to	should have told me sooner
watch out	two birds with one stone
way too	what are you up to?
white horse	<hr/> <b><i>6-word</i></b> <hr/>
would rather/ 'd rather	Are you looking down on (me/ us)?
yet another	Are you out of your mind?
	What's taking him/her so long?
	You've got to be kidding

---