



**A CORPUS-BASED STUDY OF KEYWORDS AND  
LEXICAL COLLOCATIONS USED IN TV  
TRANSCRIPTS BASED ON THE TV SERIES “YOUNG  
SHELDON”**

**BY**

**SARUNYA NOOCHAN**

**AN INDEPENDENT STUDY SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE  
OF MASTER OF ARTS IN ENGLISH LANGUAGE TEACHING  
LANGUAGE INSTITUTE  
THAMMASAT UNIVERSITY  
ACADEMIC YEAR 2023**

**A CORPUS-BASED STUDY OF KEYWORDS AND  
LEXICAL COLLOCATIONS USED IN TV  
TRANSCRIPTS BASED ON THE TV SERIES “YOUNG  
SHELDON”**

**BY**

**SARUNYA NOOCHAN**

**AN INDEPENDENT STUDY SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE  
OF MASTER OF ARTS IN ENGLISH LANGUAGE TEACHING  
LANGUAGE INSTITUTE  
THAMMASAT UNIVERSITY  
ACADEMIC YEAR 2023**

THAMMASAT UNIVERSITY

LANGUAGE INSTITUTE

AN INDEPENDENT STUDY

BY

SARUNYA NOOCHAN

ENTITLED

A CORPUS-BASED STUDY OF KEYWORDS AND LEXICAL COLLOCATIONS  
USED IN TV TRANSCRIPTS BASED ON THE TV SERIES “YOUNG  
SHELDON”

was approved as partial fulfillment of the requirements for  
the degree of Master of Arts in English Language Teaching

on July 27, 2024

Chairman

Arthitaya Narathakoon

(Arthitaya Narathakoon, Ph.D.)

Member and Advisor

Nantikarn Simasangyaporn

(Nantikarn Simasangyaporn, Ph.D.)

Director

Supakorn Phoocharoensil

(Associate Professor Supakorn Phoocharoensil, Ph.D.)

Independent Study Title	A CORPUS-BASED STUDY OF KEYWORDS AND LEXICAL COLLOCATIONS USED IN TV TRANSCRIPTS BASED ON THE TV SERIES “YOUNG SHELDON”
Author	Sarunya Noochan
Degree	Master of Arts
Major Field/Faculty/University	English Language Teaching Language Institute Thammasat University
Independent Study Advisor	Nantikarn Simasangyaporn, Ph.D.
Academic Year	2023

## ABSTRACT

This research investigates the use of keywords and their lexical collocations in the TV series “Young Sheldon” to enhance English language learning and teaching. Utilising AntConc software, the scripts of the first six seasons of “Young Sheldon” were carefully analysed to identify a list of 100 keywords and the most frequent collocations of the top 30 keywords. These keywords were systematically categorised into nouns, verbs, adjectives, and adverbs, providing a comprehensive analysis of their roles in the dialogue. The findings reveal a significant focus on action-oriented verbs, essential nouns, descriptive adjectives, and adverbs, reflecting the dynamic and varied language used in the series. By examining how these words are used together in context, the research provides useful insights into how natural conversations flow smoothly and make sense. This research highlights the importance of authentic media in language education, demonstrating how TV series can effectively enrich vocabulary, enhance understanding of natural speech patterns, and improve overall communication skills.

**Keywords:** Keywords, Collocations, Corpus-based study, Vocabulary

## ACKNOWLEDGEMENTS

I am profoundly grateful to all those who have supported me during my independent study. The guidance, support, and encouragement of numerous individuals were essential for the completion of this study.

First and foremost, I am incredibly thankful to my advisor, Dr. Nantikarn Simasangyaporn, whose insightful feedback, and constant encouragement have greatly contributed to the quality of this work.

I am grateful for all the instructors who have guided me throughout this programme. Their commitment and expertise have greatly influenced my academic path.

I would also like to thank all my peers, especially Papassara, Pattarin, and Thitiphong, for their significant assistance and continuous encouragement and for making this journey memorable and enjoyable.

Finally, my sincere gratitude goes to my family for their support during this challenging period. I am especially thankful to my partner, Ian, for his endless support, patience, and understanding throughout this journey.

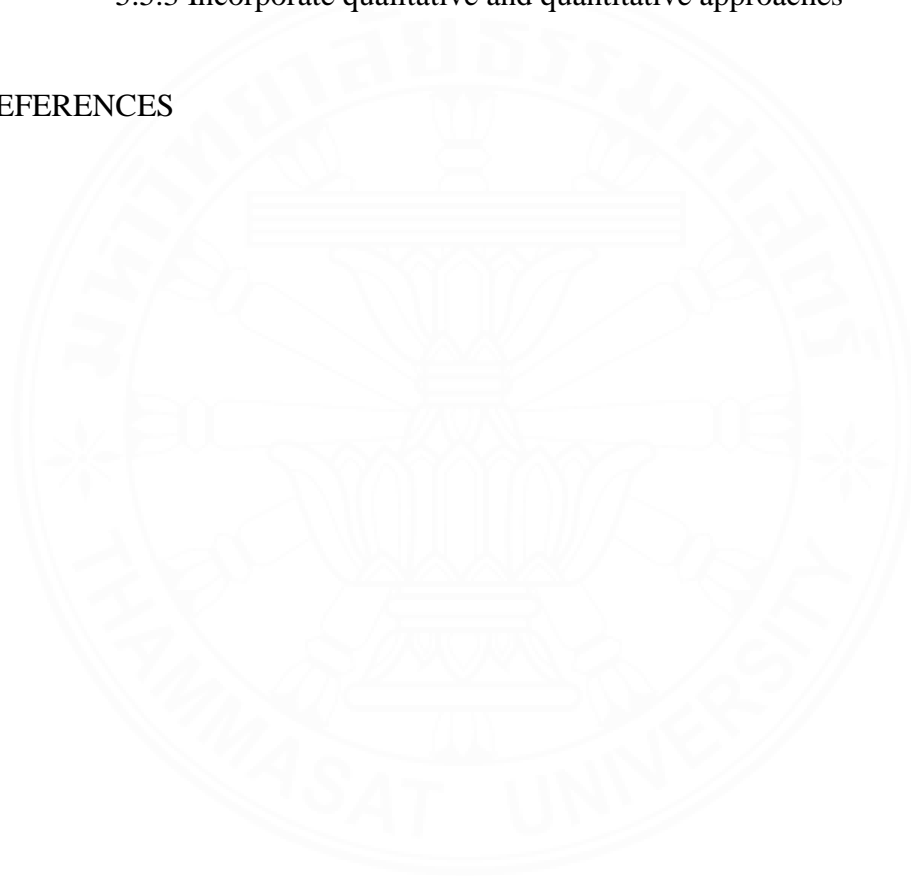
Sarunya Noochan

## TABLE OF CONTENTS

	Page
ABSTRACT	(1)
ACKNOWLEDGEMENTS	(2)
LIST OF TABLES	(6)
CHAPTER 1 INTRODUCTION	1
1.1 Background of the study	1
1.2 Statement of the problem	2
1.3 Research objectives	3
1.4 Research questions	3
1.5 Definitions of key terms	3
1.6 Scope of the study	4
1.7 Significance of the study	4
1.8 Organization of the study	5
CHAPTER 2 REVIEW OF LITERATURE	7
2.1 Corpus linguistics	7
2.2 Word frequency	8
2.3 Keywords	9
2.4 Lexical collocations	9
2.5 Previous studies	10
2.5.1 Corpus linguistics and transcripts analysis	10
2.5.2 Corpus linguistics and keyword analysis in transcript	13

CHAPTER 3 RESEARCH METHODOLOGY	16
3.1 Design of the study	16
3.2 Data source	16
3.2.1 AntConc	16
3.2.2 The Young Sheldon series corpus	16
3.2.3 AME06 corpus	18
3.3 Data procedure	18
3.3.1 Identifying keywords	18
3.3.2 Identifying collocations of the keywords	19
CHAPTER 4 RESULTS	21
4.1 The list of the top 100 keywords	21
4.2 The most frequent collocations of the top 30 keywords	24
4.2.1. Keywords and their collocations related to family and people	26
4.2.2. Keywords and their collocations related to religion	26
4.2.3. Keywords and their collocations related to common daily conversation	27
4.2.4. Keywords and their collocations related to feelings or opinions	29
4.2.4.1 Feelings	29
4.2.4.2 Opinions	31
4.3 Collocation types and combinations	31
CHAPTER 5 DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS	34
5.1 Summary of the study	34
5.2 Summary of the findings	34

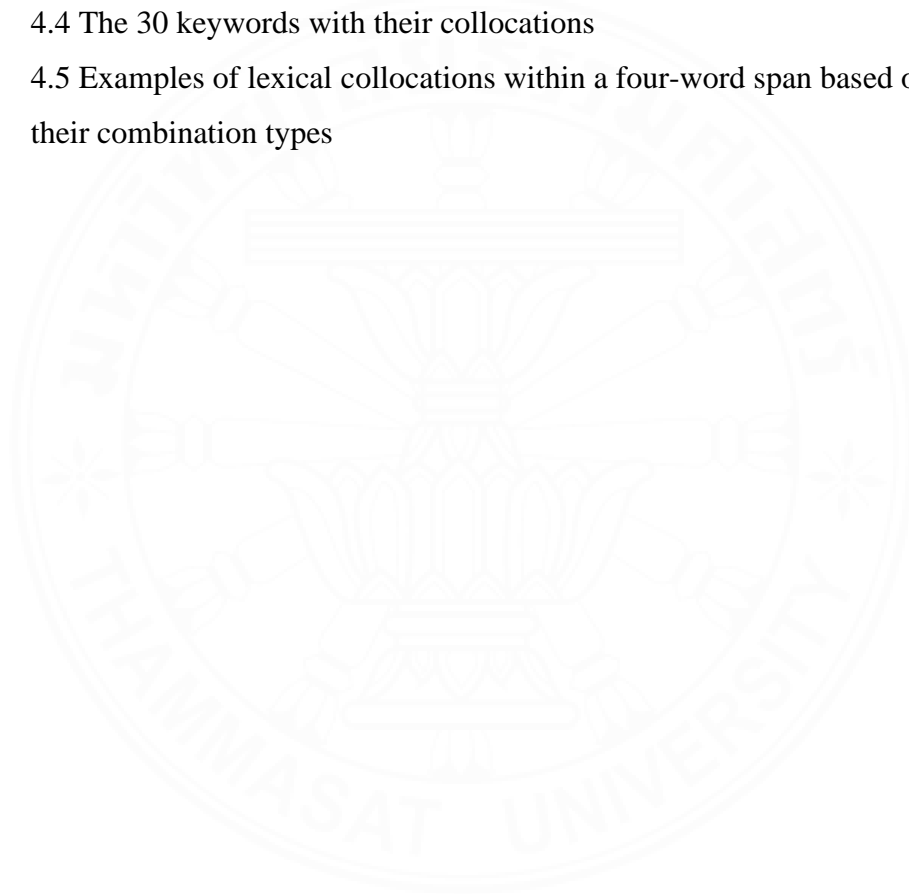
	(5)
5.2.1 The 100 keywords	35
5.2.2 The collocations of the top 30 keywords	36
5.3 Implications	38
5.4 Limitations of the study	39
5.5 Recommendations	39
5.5.1 Broader media analysis	39
5.5.2 Expanded focus	40
5.5.3 Incorporate qualitative and quantitative approaches	40
REFERENCES	41





## LIST OF TABLES

Tables	Page
4.1 The top 100 keywords in the Young Sheldon series corpus	21
4.2 The top 100 keywords in six main categories based on their meanings	23
4.3 The top 100 keywords in four grammatical categories	23
4.4 The 30 keywords with their collocations	24
4.5 Examples of lexical collocations within a four-word span based on their combination types	32



# CHAPTER 1

## INTRODUCTION

### 1.1 Background of the Study

It is essential for those learning a foreign language to consistently increase their knowledge of new and more complex words. By expanding their vocabulary, they can communicate with greater precision and comprehension. In addition, it enables them to convey complex ideas and emotions effectively. The significance of words cannot be underestimated. They can activate a student's prior knowledge and experiences, allowing them to construct meaning (Johns & Wilke, 2018).

According to Lewis (2000), a student with a vocabulary of 2,000 words can only communicate in a limited way, while a student who has mastered collocations with the same vocabulary of 2,000 words will be significantly more effective in communication. Collocation, the frequent occurrence of certain words together in a language, is a key aspect. Understanding how a word combines with other words in actual use, such as collocations, is just as important in determining its meaning as the meaning it holds on its own (O'Keeffe et al., 2010). McCarthy and O'Dell (2005) emphasised that collocations give importance to pairs of words as integrated chunks of meaning and usage rather than individual words. Mastering collocations enhances your language skills and makes your communication more natural and fluent.

Nowadays, collocation has become a widely accepted aspect of vocabulary description and pedagogy (as cited in O'Keeffe et al., 2010). Mastering the collocations of a second or foreign language is crucial for anyone who wishes to achieve more than a basic level of proficiency in the language, as collocation is present even in the most common and fundamental words (O'Keeffe et al., 2010). Instead of focusing on each individual word, learners can use collocations to produce language more automatically and confidently. By understanding collocations, learners can improve their comprehension skills as it allows them to recognise common word combinations and comprehend their meanings in context. To ensure that learners of a target language become proficient in using it, it is crucial for language teachers to offer instructions on collocation (Sukman et al., 2022).

Utilising TV series as instructional materials offers a unique and enjoyable approach to expanding learners' vocabulary and language skills. TV series serves as a valuable resource for language learners, offering authentic and contextualised language usage that exposes learners to real-life conversations. This is one of the most effective ways to acquire a language. Watching TV shows or series in the target language is a popular pastime among learners, and it can assist in enhancing one's speaking abilities by enriching both listening skills and active vocabulary (Sezgin & Öztürk, 2020).

TV series provides a unique learning environment where words and phrases are used in real-life situations, allowing learners to grasp their contextual meaning. Firth (1935) emphasised that the full significance of a word is constantly dependent on its context, and any examination of meaning without considering the complete context cannot be considered valid. Furthermore, TV series often feature recurring themes, characters, and situations, so learners are repeatedly exposed to the same vocabulary. This repetition can reinforce learning and aid in memorising. Watching TV series also allows learners to enhance their language skills in an enjoyable and captivating way.

## **1.2 Statement of the Problem**

TV series can be an excellent tool for teachers to enhance their students' vocabulary and improve their language use in real-life situations. This tool improves language proficiency and provides a deeper understanding of the culture (Lengerlioğlu, 2019). However, despite its potential, there has been little investigation into the relationship between keywords and their collocations from TV shows and vocabulary teaching which could be valuable material for teachers to use. Additionally, this kind of study can make a contribution to the existing field of corpus linguistics.

One popular TV series that is definitely worth investigating and using as instructional material is "Young Sheldon." This TV series is broadcast on CBS in the USA and is also available on a subscription-based streaming service, Netflix. Miščin (2018) pointed out that television is a common form of media in our era. We have been familiar with it since our early years, and it is often one of the first ways in which we encounter a different language, whether through TV programs, animated series, or YouTube videos. Teachers can utilise "Young Sheldon" to improve their students' language skills and provide them with a fun and engaging learning experience.

### **1.3 Research Objectives**

- 1) To create a list of 100 keywords found in TV transcripts based on The TV series “Young Sheldon”.
- 2) To explore the most frequent collocations of the top 30 keywords found in TV transcripts based on The TV series “Young Sheldon”.

### **1.4 Research Questions**

- 1) What is a list of 100 keywords found in TV transcripts based on The TV series “Young Sheldon”?
- 2) What are the most frequent collocations of the top 30 keywords found in TV transcripts based on The TV series “Young Sheldon”?

### **1.5 Definitions of Key Terms**

#### ***1.5.1 Young Sheldon Series Corpus***

It is a custom-made corpus consisting of transcripts from all six seasons of the TV show Young Sheldon, which was compiled specifically for this study. This corpus serves as the primary data source for the study.

#### ***1.5.2 AME06 Corpus***

It is a corpus comprised of one million words of published American English writing. It includes 500 files, each containing 2000-word samples taken from 15 different genres of writing.

#### ***1.5.3 Keywords***

In this study, keywords are significant content words identified from the Young Sheldon series corpus, excluding function words. These are words like nouns, verbs, adjectives, and adverbs that appear unusually frequently in the Young Sheldon series corpus compared to the AME06 corpus.

#### **1.5.4 Collocations**

Collocations refer to pairs or groups of words that frequently appear together within a specific span of text. This study focuses on identifying and analysing the collocations of the top 30 keywords within a four-word span on either side of each keyword.

#### **1.6 Scope of the Study**

This study aims to examine a list of 100 keywords and their lexical collocations used in the dialogues of the TV series “Young Sheldon.” These transcripts were compiled using Antconc software to create a Young Sheldon Series Corpus. The AME06 corpus, created by Paul Baker, serves as a reference file or benchmark corpus for comparison with the Young Sheldon Series Corpus to identify keywords from Young Sheldon transcripts. However, function words have been excluded from the analysis, so the 100 keywords are all content words, such as nouns, verbs, adjectives, and adverbs.

“Young Sheldon” is a prequel sitcom to “The Big Bang Theory”. It was created by Chuck Lorre and Steven Molaro in 2017. The show focuses on the life of Sheldon Cooper, a gifted child who grew up in East Texas during the late 1980s and early 1990s. The program covers a range of topics, including science, family relationships, school life, and social interactions. It explores a variety of themes, including family relationships, intellectual curiosity, social awkwardness, and the challenges of fitting in. The primary focus of the dialogue in the show is centred on the interactions that Sheldon has with his family members, such as his mother, father, brother, twin sister, and grandmother. These interactions are the central theme of the show’s dialogue. Therefore, the dialogue mainly focuses on everyday language with a Southern American accent, except for the main character, Sheldon, who always speaks formally and uses technical scientific language.

#### **1.7 Significance of the Study**

Based on a corpus methodology, this study was conducted to identify 100 keywords and their collocations, as well as to examine the difficulty level of these keywords. The data was extracted from the transcripts of all six seasons of the TV show

“Young Sheldon”. The findings from this study can be a valuable resource for English Language Teaching (ELT). Teachers can utilise the outcomes of this study in their teaching materials to enhance students’ vocabulary. Moreover, teachers can guide students on how to use these terms appropriately in different contexts. Instead of focusing on individual words, language learners can use collocations to produce speech more naturally and confidently. By understanding collocations, learners can improve their comprehension abilities by recognising typical word combinations and comprehending their contextual meanings. Teachers can also design tasks or exercises based on the findings to evaluate students’ comprehension, vocabulary mastery, and ability to use and understand vocabulary in context.

The study can serve as a model for corpus-based language teaching and learning research. Similar methodologies can be adopted to analyse language usage in other texts or TV series for pedagogical implications. In summary, analysing the keywords and lexical collocations in TV transcripts, specifically from a well-known show such as “Young Sheldon,” can be extremely useful for teachers seeking to improve language instruction, contextual vocabulary acquisition, cultural awareness, and language proficiency development for their students.

### **1.8 Organization of the Study**

This study is divided into five chapters, listed as follows:

1) Chapter One serves as the introduction to the study. It provides the background of the research and the statement of the problem and outlines the research objectives and research questions. The chapter also defines key terms, explains the scope of the study and significance of the study, and discusses the organisation of the research.

2) Chapter Two provides a literature review of corpus linguistics, including word frequency, keywords, and lexical collocations, followed by a discussion of previous studies.

3) Chapter Three focuses on the methodology used to conduct the research.

4) Chapter Four reports the findings and discussions of the research.

5) Chapter Five includes a summary of the study conducted, followed by a summary of the findings. The chapter concludes with a conclusion and recommendations for future research.



## **CHAPTER 2**

### **REVIEW OF LITERATURE**

This chapter is divided into five different sections. The first part provides an overview of corpus linguistics, while the second part delves deeper into the concept of word frequency. The third part explains the concept of keywords, followed by a discussion on lexical collocations in the fourth part. Lastly, the fifth part presents previous relevant research studies.

#### **2.1 Corpus Linguistics**

The field of corpus linguistics emerged in recent times as a subfield of linguistics. It has gained popularity, particularly since the introduction of personal computers in the 1990s (Baker, 2010). A corpus is a methodical gathering of spoken or written text, which is turned into computer-stored collections (O’Keeffe et al., 2010). These texts are suitable for qualitative and quantitative analysis as they represent real-life examples of their usage (McEnery & Wilson, 1996, as cited in Baker, 2010). The term corpus originates from the Latin word for body; its plural form is corpora. Essentially, a corpus is a collection of language, often a very extensive one, that has occurred naturally and is saved as computer files (Baker, 2010).

As mentioned, language features in a corpus can be analysed using quantitative and qualitative approaches. The quantitative approach involves counting the occurrences of specific words in a corpus and comparing them with the frequencies in other corpora. On the other hand, the qualitative approach involves examining how words are used across a corpus and looking beyond the frequency of word occurrences. This approach can be achieved by analysing concordance lines to identify qualitative patterns of use that go beyond frequency (O’Keeffe et al., 2010).

Moreover, the use of corpora in language studies is beneficial because they cover a wide range of language usage, making them more representative of real-life language than smaller datasets or intuitive judgments. This enables more reliable generalisations about language. A corpus is a large collection of texts that is carefully designed to be representative of a particular language or language variety. It serves as



a sample that can help in studying language patterns and trends. Corpora are constructed with great care to ensure that they reflect the linguistic features of the language or language variety being studied (Baker, 2010).

Furthermore, there are various techniques that can be applied to a corpus (O’Keeffe et al., 2010). One of these methods is the concordance technique, which uses the corpus software to locate every instance of a specific word or phrase. Another technique is the rapid calculation of word frequency lists, allowing for easy comparison across different corpora, language varieties, and contexts. Additionally, there is a keyword analysis function that can identify the most important words in one or more texts.

## **2.2 Word Frequency**

Another frequently used technique in corpus analysis involves word frequency counts. Corpus analysis software can rapidly generate word frequency lists (or wordlists) for any given set of texts. This word frequency feature can be performed on a corpus to obtain a list of all the words in a corpus ranked according to their frequency. This feature enables analysis across various corpora and diverse contexts of use (O’Keeffe et al., 2010). Information about word frequency in a corpus provides insights into the number of words that native speakers use, the frequency of the individual words they use, and how they combine those words (O’Keeffe et al., 2010, as cited in Timmis, 2015). Furthermore, in corpus linguistics, word frequency analysis has shown that most of the language we encounter consists of a small number of highly common words. This discovery has important implications for the field of English Language Teaching (ELT) (Timmis, 2015).

For example, according to Nation (2013) as cited in Timmis (2015), around 80-90% of the language used in conversations or books can be attributed to the 2,000 most frequently used words in English. Incorporating a list of the most common words into vocabulary teaching can be highly beneficial, as these words are crucial for communication in various situations. Learners can acquire useful vocabulary that they will often come across and use in their daily conversations and written communications by learning these words. English language teachers can utilise word frequency analysis to initiate the creation of a vocabulary syllabus (Timmis, 2015).

### **2.3 Keywords**

In corpus linguistics, the most frequent words in a text or group of texts are not typically considered keywords. Instead, keywords are those that appear with an unusual frequency (O'Keeffe et al.,2010). According to Rayson (2019), a keyword is a word that is examined for its importance or significance when comparing two corpora and it is a word that holds significance and is used to describe the content of a document. In the field of corpus analysis, the definition of a "keyword" expands to a word that is statistically significant and characterises a document, text, or corpus. Keywords have a quantitative value as they are words or terms that are statistically distinctive in a text or a group of texts (Biber & Reppen, 2015). Kyröläinen and Laippala affirmed that in corpus linguistics, a keyword is a word or phrase that is utilised to examine and recognise the attributes of a discourse domain that is linked with texts. Keywords hold importance in two related senses: Firstly, they have a crucial role as binding words in specific activities and their interpretation. Secondly, they are also significant as indicative words in particular forms of thought (Raymond, 2014). The importance of keywords is determined by their representation and influence on culturally significant discourses. According to Egbert and Biber (2019), keywords are closely linked to the subject matter of texts within a specific discourse domain (as cited in Szczygłowska, 2022).

### **2.4 Lexical Collocations**

Lexical collocations are groups of words that frequently appear together more often than expected by chance. These word combinations are not random and occur frequently within a language. Essentially, lexical collocations refer to the co-occurrence of words or lexical items within a corpus of text. Collocations, as described by Conzett et al. (2000), are words that frequently appear together. These words can be as short as two words, like “problem child”, or as extended phrases like “He is recovering from a major surgery.” Collocations are commonly used in spoken and written communication, and they form a pattern in language. The basis for these collocations is statistical observation, which involves analysing how often certain words appear in proximity to each other within a corpus of text. According to Tutin (2009), collocations were defined by Halliday and Hasan (1978) as lexical elements that occur repeatedly

and contribute to the coherence of a text. According to Hausmann's 2004 study (as cited in Tutin, 2009), one component of a collocation, which is referred to as the “base,” retains its typical definition as a word with inherent meaning (also known as autosemantic words). Meanwhile, the other component, called the “collocate,” relies on the base word (synsemantic words) and generally has a less straightforward definition.

## **2.5 Previous Studies**

The following section reviews important studies that use corpus linguistics, transcript analysis, and keyword analysis to analyse language use in media. These studies provide valuable insights and applications for language learning, demonstrating how media can enhance language understanding and skills.

### ***2.5.1 Corpus Linguistics and Transcripts Analysis***

Corpus linguistics and transcript analysis are crucial for understanding language use in media. This section reviews key studies using these methods. I.S.P. Nation (2006) identified the minimum vocabulary needed for comprehension across various media. Sezgin and Öztürk (2020) showed that TV series language aligns with real-life spoken language and enhances language skills. Wahyuni (2021) examined cohesion in "The Lion King" script, highlighting the role of cohesive devices. Patumanon (2016) studied intensifiers in "Friends," comparing them with real-life conversations. These studies provide valuable insights and applications for language education.

To elaborate further, I.S.P. Nation (2006) conducted research to determine the minimum size of receptive vocabulary required for common language activities such as reading a novel, perusing a newspaper, watching a film, and engaging in a conversation. The researcher stated that it is crucial to establish the minimum vocabulary required to understand a text or movie or the threshold of unfamiliar words allowed in a text before it starts to impede comprehension. The order in which vocabulary is learned tends to follow its frequency and range, regardless of whether one is a native or non-native speaker. Typically, words that are used more often and have a broader range of meanings are learned before less common and more specific words. If a person's reading or viewing material aligns with their ability level, they will only come across around 1% of unfamiliar words. This means that the first list of 1,000 words

should contain more families, tokens, and types than the subsequent lists in the same material. In this study, the researchers analysed the script of the popular children's movie *Shrek* and found that it has around 10,000 words and uses 1,100 different word families. The majority of the words in *Shrek* are derived from the first 1,000 most commonly used words in the English language, and on average, there is one unfamiliar word encountered for every 30 words. Nevertheless, the film's strong visual elements allow children of all ages to enjoy and even memorise the script.

Expanding on the idea of integrating language learning with authentic spoken language, Sezgin and Öztürk (2020) conducted a study to analyse how much real spoken language is used in TV series in terms of vocabulary. For this study, a British TV Series Corpus (BTSC) corpus was compiled using two British TV series, *Sherlock* and *Doctor Who*. The spoken part of the British National Corpus (BNC) was compared to this corpus. The study had three main objectives. Firstly, the researchers wanted to explore the extent to which the BTSC contains the items that are present in the BNC spoken frequency lists. Secondly, they aimed to establish a significant correlation between the frequency of the items in the spoken part of the BNC and the BTSC. Lastly, their aim was to identify similarities between the BNC and the BTSC concerning the 20 most frequent non-lemmatized and lemmatised items.

The frequency list used for data collection in the spoken section of the BNC includes only items that have a minimum frequency of 10 per million words. This means that words with frequencies lower than this are not included in the present study's lists. This condition applies to the entire corpus list, which has a minimum frequency of 160 per million words. Data for the British Television Series Corpus (BTSC) was collected by sourcing scripts from the official BBC website in .docx format. The word count for *Sherlock* and *Doctor Who* was 120769 and 633609 respectively. The Textworks 1.5.6 software was used for analysis and to produce frequency lists after merging the scripts. Proper names, misspelt words, contractions, abbreviations, and exclamations were manually excluded from the BTSC. The findings showed that TV series reflect real-life language and can improve vocabulary, speaking, and listening skills. They cover 98.54% of frequent lemmas in the British National Corpus and can make effective teaching materials for in-class and extra-curricular settings. The paired samples t-test was used to compare the BNC and the BTSC in terms of item frequency

on SPSS. The result showed that there was no significant difference between the frequency of the items on the BTSC and the BNC lists.

It has been observed that corpus linguistics is valuable not only for analysing the vocabulary of real spoken language used in TV series but also for examining the lexical and grammatical cohesion present in a film script. Recently, Wahyuni (2021) conducted a study on *The Lion King Movie Script*, analysing the grammatical and lexical cohesive devices used in the script. The researcher focused on identifying the types of grammatical and lexical cohesive devices used, describing how they were used, and investigating their overall impact on the script. The results revealed that there were 312 ties of grammatical cohesive devices and 196 ties of lexical cohesive devices. The cohesive devices used in the script had a semantic concept that fell between lexical and grammatical cohesion devices, indicating relationships within and outside of the text. The most frequently used types of grammatical cohesion were personal reference and demonstrative reference. However, clausal substitution did not occur in the whole scene. Lastly, the most used types of lexical cohesion were repetition. This device was used to rewrite an item in a preceding element with the same form in the following element.

Another linguistics feature has been studied in a TV series by Patumanon (2016). The researcher explored the use of five intensifiers - so, really, very, too, and pretty - aiming to compare their usage in the TV series *Friends* and real-life conversations, focusing on their distribution, adjectival collocations, and semantic prosody. The researcher analysed *Friends*' transcripts and selected spoken language registers from COCA for investigation. They explored the top 20 adjectives of each intensifier and conducted distributional and collocation analyses. Semantic prosody data were grouped into categories for comparison. They found that "so" was the most frequently used intensifier in *Friends*, while "very" was the most preferred in COCA. The adjectival collocations of "so" and "really" were similar in both corpora. "Too" was the most consistent according to the semantic prosody analysis. Very occurs more frequently with positive and negative adjectives in COCA, while so and really occur more frequently with positive and negative adjectives in *Friends*.

Overall, these studies collectively contribute to a comprehensive understanding of how corpus linguistics and transcript analysis can be utilised to explore and analyse

language use in media. They provide methodological frameworks and insights that are directly applicable to the present study, highlighting the importance of essential vocabulary, real-life language use, lexical and grammatical cohesion, and the use of intensifiers in media transcripts.

### ***2.5.2 Corpus Linguistics and Keyword Analysis in Transcript***

Keyword analysis in corpus linguistics reveals the frequency and importance of words in specific contexts. This section reviews key studies using this method in various media. Hamed (2021) analyzed the discourse of US presidents, highlighting language differences. Scheffler, Jones, and Domińska (2021) examined “Peppa Pig,” showing its language supports learning. Siripachanah (2022) studied “Desperate Housewives,” identifying engaging thematic keywords. These studies demonstrate the value of keyword analysis. This section sets the stage for examining “Young Sheldon,” focusing on everyday language and cultural references to enhance language learning.

To begin with, Hamed (2021) utilised corpus-assisted analysis of keywords to investigate the discourse of the past four US presidents, focusing on the meanings of the keywords and their collocations. Hamed (2021) explained that salient words in a corpus that occur more frequently than usual (positive keywords) or less frequently than usual (negative keywords) when compared to a reference corpus are called keywords. The co-occurrence of words is known as collocation. The researcher employed AntConc, a software for processing corpora, to generate the investigation of keywords and collocations. The findings showed the similarities and differences between the four most recent American Presidents regarding their key topics.

The study revealed the different priorities and language used by Democratic and Republican Presidents in the United States. It compares the social and foreign policy focuses of Presidents Clinton and Obama, highlighting Obama's greater emphasis on foreign affairs. It also contrasts President Bush's focus on external issues and the war on terrorism with his negative language. Additionally, the study unveiled President Trump's language, noting the exaggerated positive descriptions and negative keywords related to program reform and external issues.

Keyword analysis has also been applied in the context of children's media, which is valuable for using it as vocabulary input for young learners. For instance,



Scheffler, Jones, and Domińska (2021) conducted research by compiling a corpus from the well-known animated show "Peppa Pig." The study aimed to compare the most common words, keywords, and formulaic sequences used in the show to those found in the spoken section of the British National Corpus (BNC). The researchers found that the language used in this series reflects the world of the characters, using keywords such as hamster, gazelle, and blackberry. The child-like language is focused on topics such as animals, food, and weather which are of immediate interest to young viewers. The use of the word "hello" is also prominent in the children's language used in this series, serving as a narrative device to help young viewers understand who is talking to whom and to orient themselves to the program. The intention is to model positive social interactions and portray characters who are friendly to each other, making it potentially useful input for pre-primary English as foreign language learners and language acquisition in general, according to the analysis of the corpus data.

Furthermore, keyword analysis and corpus linguistics have been employed to examine the styles of TV transcripts, as can be seen in Siripachanah's (2022) study on the style of one of the most popular American TV shows, *Desperate Housewives*. They analysed the keywords present in the show's scripts and compared them to a corpus of American TV series that were broadcast during a similar time period to the target corpus. The researcher gathered scripts of all episodes from various online platforms, including websites and blogs, to create a corpus named *Desperate Housewives Corpus* (DHC). The DHC consists of scripts from all eight seasons of the TV series *Desperate Housewives*. The researcher used quantitative methods to identify keywords in the popular TV series. All the keywords were sorted into four primary categories, namely, fictional world signals, thematic signals, spoken language, and others, based on an analysis of their meanings and concordance lines. The four major groups were each further broken down into specific subgroups based on the keywords. Three thematic-signal keyword subgroups were chosen for qualitative analysis: food and beverages, family and relationship, crime and mystery. The researcher found that the three groups of keywords play a significant role in portraying the four housewife protagonists' personalities and various themes covered in the series, such as the pressure on women's domestic roles and female unity. Additionally, some keywords, particularly those related to crime and mystery, contribute to the show's lasting popularity among viewers

due to its ability to maintain their curiosity and engagement with the series. Siripachanah's study offers insights into the importance of keywords in portraying characters and themes in media and highlights the role of certain keywords in sustaining audience engagement.

Reviewing previous studies on corpus linguistics and keyword analysis in transcripts shows that the keywords found vary in each study due to the nature of the TV shows and the context in which the keywords are represented. Hamed's (2021) analysis of presidential discourse highlights the importance of identifying and understanding keyword usage and their collocations to reveal underlying themes and priorities. This approach is directly applicable to examining how "Young Sheldon" employs keywords and collocations to convey character dynamics and plot development.

Similarly, the study by Scheffler, Jones, and Domińska (2021) on "Peppa Pig" highlights the value of children's media for language learning, aligning with the aim of this study to use "Young Sheldon" as a resource for teaching everyday language and cultural references

Furthermore, Siripachanah's (2022) research on "Desperate Housewives" shows how keyword analysis can reveal stylistic elements and audience engagement. This supports the goal of enhancing language learning through authentic, relatable content.

These previous studies demonstrate how examining vocabulary, some linguistic features, keyword usage, and collocations in various media contexts can reveal important linguistic patterns and themes. By analysing "Young Sheldon" in a similar manner, this research aims to enhance language learning through authentic and relatable content, making it a valuable resource for teaching vocabulary in context through everyday language and cultural references. The insights from these studies support the goal of using TV series as effective tools for language teaching.



## **CHAPTER 3**

### **RESEARCH METHODOLOGY**

In this chapter, I will discuss the methodology of data collection in three sections. The first section, 3.1, will focus on the study's design, while the second section, 3.2, will discuss the study's data source. The third section, 3.3, will discuss the data procedures used to identify keywords and their collocations.

#### **3.1 Design of the Study**

The present study is a corpus-based study which aims to identify the top 100 keywords from the Young Sheldon Series Corpus and determine the most frequent collocations of the top 30 keywords found in the corpus. The Young Sheldon Series Corpus was created from the scripts of the American TV show "Young Sheldon" using Antconc software to compile all the scripts from the first 6 seasons. Using Antconc's keyword and collocate tools, 100 keywords and the collocations of the top 30 keywords were generated from the script. The collocations were further examined for their contextual meaning based on the language used in the show.

#### **3.2 Data Source**

##### **3.2.1 *AntConc***

Anthony (2005) described AntConc as a toolkit for corpus analysis that could be utilized in classrooms. It is free, multi-platform, and multi-purpose. It comes with a powerful concordancer, tools for cluster and lexical bundle analysis, a word distribution plot, and generators for word and keyword frequency. AntConc is a great option for individuals, schools, or colleges with a limited budget as it is a freeware application that can work on Windows, MacOS, and Linux/Unix-based systems. Though it is freeware, it is equipped with an easy-to-use interface and provides advanced features for corpus analysis.

##### **3.2.2 *The Young Sheldon Series Corpus***

The Young Sheldon TV show has been on air since 2017 and has a total of 130 episodes spread across 7 seasons. Each episode runs for approximately 20 minutes. The

Young Sheldon Series Corpus is a collection of scripts from all episodes of the show's first six seasons in Microsoft Office Word format (.docx). These scripts were obtained from the internet via <https://transcripts.foreverdreaming.org> (2024). Unfortunately, since the seventh season's episodes have not yet been aired at the time of compiling the corpus, the Young Sheldon Series Corpus does not include any scripts from the seventh season. Therefore, the Young Sheldon Series Corpus includes only 127 episodes from the first six seasons. After compiling transcripts from all episodes in the first six seasons, the data was imported into AntConc, a software tool used for corpus analysis.

To ensure the accuracy and completeness of the transcripts, I personally re-watched all the episodes from seasons 1 to 6. This meticulous process allowed me to cross-check the scripts with the actual dialogues, ensuring that the Young Sheldon Series Corpus is a reliable and accurate representation of the language used in the show. By re-watching the episodes and comparing them with the transcripts, I was able to validate the accuracy of the corpus data, thus enhancing the credibility of the research findings.

The TV show "Young Sheldon" is a suitable choice for this study, as it provides a vast amount of language data that includes dialogue from several characters in varying contexts. These contexts include science, family relationships, school life, and social interactions, which enable the study to analyse and identify numerous keywords and their collocations across various settings. Educators can use the findings to create useful lessons that challenge and enhance students' vocabulary and place it in different contexts. Additionally, teachers who aim to teach everyday language can benefit from understanding and learning the keywords and their collocations from this show. The series portrays everyday interactions between characters, including family members, classmates, and neighbours. These conversations usually centre around common topics such as school, family life, friendships, and hobbies, which provide learners with exposure to authentic, natural language usage in familiar contexts.

The TV show "Young Sheldon" may not be the best representation of the language Thai learners should learn. This is because the show primarily features English dialogue that is set in a specific cultural and linguistic context. The show is set in East Texas during the late 1980s and early 1990s and most of the characters in the series speak with a southern American accent. However, the show does offer insights

into American culture, society, and daily life during that time period. By understanding cultural nuances, different accents in English, and new and more challenging vocabulary, learners can improve their overall language proficiency and enhance cross-cultural communication.

### ***3.2.3 AME06 Corpus***

The AmE06 Corpus is a compilation of American English, which includes one million words. It was formulated by Paul Baker, who is a professor of Linguistics and English Language at Lancaster University (Baker, 2007). The corpus contains 500 files, each with 2000-word samples extracted from 15 different genres of writing. The majority of the texts in the corpus were published in 2006. This wordlist corpus can be loaded into AntConc (version 4.0 or later) using the Corpus Manager menu and can be used as a reference corpus to generate keyword lists (Anthony, n.d.).

This corpus was chosen as the reference corpus because it is a large collection of American English, just like the language used in the “Young Sheldon” series. Additionally, the AME06 Corpus is bigger than the “Young Sheldon” series corpus, making it a strong benchmark for generating keyword lists. These reasons make it suitable for this study.

## **3.3 Data Procedure**

### ***3.3.1 Identifying Keywords***

The next step was to generate a list of keywords in the Young Sheldon series Corpus, which is the target corpus being analysed in this study. This was done by comparing the word list in Young Sheldon with a reference corpus or benchmark corpus called AME06. This method was done by using the keyword list tool in Antconc to calculate the ‘keyness’ of words using log likelihood statistical measures in order to identify keywords, which are words that occur statistically unusually frequently. The keyword list tool is designed to identify words that occur unusually frequently in the target corpus compared with the reference corpus. This comparison is based on the same words in both corpora (Anthony, 2004).

The keyword feature setting used in this study is Text Dispersion Keyness (4-term). This setting does not simply look at the frequency of words in the target corpus

compared to the reference corpus; instead, it looks at the number of files in which they appear in each corpus. As a result, if a word appears in an unusually frequent number of files in the target corpus compared to the reference corpus, it will have a high keyness value. The higher the keyness value, the more unique it is to the Young Sheldon Series Corpus. Furthermore, the threshold setting was set at  $p < 0.05$  (3.84 with Bonferroni), according to Anthony (2005). This enabled the program to generate keywords that reflect the unusually frequent words from the Young Sheldon series corpus.

After extracting the keywords from the Young Sheldon series corpus using the keyword tool in AntConc software, I manually removed any characters' names and other function words that might have remained in the extracted keywords. This ensured that all the keywords from the Young Sheldon series corpus are content words, providing a more accurate representation of the vocabulary used in the show.

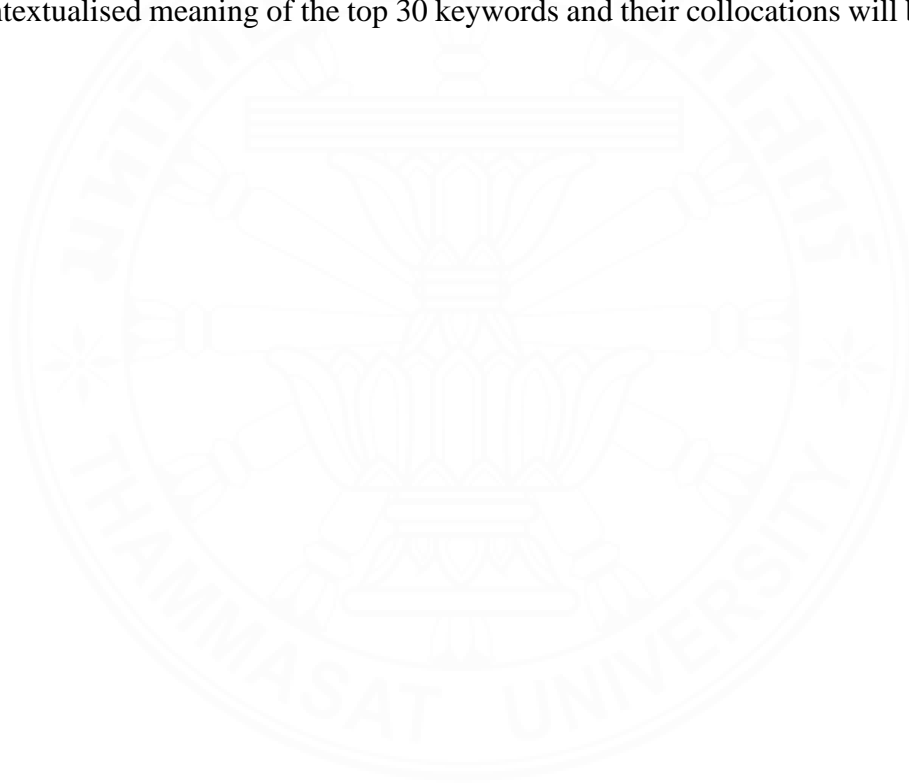
### ***3.3.2 Identifying Collocations of the Keywords***

Using keywords as “nodes” was part of the process to identify their collocates. The tool used for generating a list of common collocates of each keyword within a four-word span on the left and right of the node was the collocate tool in AntConc. This approach facilitated the exploration of collocates that occur before or after the nodes within a reasonable range of word span that is neither too close nor too far for each pair to co-occur. There are two statistics used to determine if nearby words collocate or not. The first is a probability measure, which is a log-likelihood measure. The log-likelihood measure is a statistical measure used to rank collocations based on how closely they appear statistically compared to chance. A statistical threshold used for the log-likelihood value was  $p < 0.05$  (3.84 with Bonferroni), so all log-likelihood values above 3.84 cleared the set threshold. The second is the effect size measure in this study, which is MI or mutual information, which is another statistical measure used to rank the strength of the collocation.

Only lexical collocations with a frequency of occurrence of 10 or more were selected. Additionally, collocations were chosen if their Mutual Information (MI) scores were 3 or higher. Hunston (2022) suggests that a Mutual Information (MI) score of 3 or above indicates a strong enough association to consider two words as collocates. This score indicates a statistically significant connection, suggesting that the words

often appear together within a particular context. If no collocates met the criteria of occurrences  $\geq 10$  and an MI value  $\geq 3$ , the criteria could be adjusted as needed. As a result, combinations containing pronouns, conjunctions, the verb 'to be' form, and proper nouns were removed since they were not the focus of this work. Finally, the collocations for the keywords were identified.

The next step is to display the top 100 keywords found in the YSSC corpus in a table and categorise them into four grammatical classes. They will then be grouped into main areas based on their meaning. Additionally, the most frequent collocations of the top 30 keywords will be shown in the table. Furthermore, the interpretation of the contextualised meaning of the top 30 keywords and their collocations will be provided.



## CHAPTER 4

### RESULTS

This chapter provides a detailed presentation and analysis of the study's findings. Its purpose is to thoroughly examine and interpret the data gathered during the research process using Antconc software. This analysis aims to identify keywords and their collocations, ultimately providing insights into the research questions and objectives outlined in Chapter 1.

#### 4.1 The List of the Top 100 Keywords

The corpus-based study has revealed a long list of the top 100 keywords, which are words that statistically occur with the highest frequency and keyness in the Young Sheldon series corpus. Furthermore, Table 1 on the next page offers a reflection of the language utilised by the characters. In the dataset of the Young Sheldon series, a significant portion of the top 100 keywords is centred on various aspects of the everyday life of a boy from when he was 9 years old in Season 1 until he was 13 years old in Season 6.

**Table 1**

*The Top 100 Keywords in the Young Sheldon Series Corpus*

No.	Freq.	Keyness	Keywords	No.	Freq.	Keyness	Keywords
1	79	322.714	sorry	51	411	97.131	great
2	76	309.319	nice	52	245	96.768	believe
3	111	261.717	fine	53	100	96.528	wants
4	66	250.749	fun	54	520	95.364	school
5	288	214.536	sure	55	50	88.824	football
6	306	207.605	tell	56	194	88.767	problem
7	51	188.538	stuff	57	820	87.902	see
8	200	184.845	talk	58	254	87.71	mind
9	359	184.49	really	59	109	84.798	brother
10	115	179.5	wrong	60	145	83.983	phone
11	174	179.221	boy	61	386	83.344	mother
12	62	168.088	kid	62	124	82.881	guy
13	622	159.361	want	63	115	81.612	ready
14	90	151.596	wait	64	171	81.546	understand
15	122	151.28	pretty	65	358	80.008	young
16	96	149.053	dinner	66	362	80.008	night
17	576	147.704	go	67	106	77.809	sweet
18	347	146.539	love	68	599	77.417	take
19	138	146.284	baby	69	187	76.522	son

No.	Freq.	Keyness	Keywords	No.	Freq.	Keyness	Keywords
20	66	145.881	sounds	70	68	75.773	sit
21	85	144.619	hell	71	241	74.729	keep
22	61	144.619	worry	72	227	73.915	science
23	588	143.669	think	73	265	73.472	call
24	51	143.247	cool	74	297	73.271	door
25	51	137.982	eat	75	195	72.722	stop
26	226	135.112	bad	76	627	71.38	never
27	604	131.939	say	77	531	70.167	home
28	422	127.226	look	78	408	70.075	god
29	670	126.647	here	79	244	67.152	care
30	155	125.896	ask	80	331	67.055	big
31	56	125.516	previously	81	52	67.038	interesting
32	479	120.897	need	82	154	65.94	happened
33	607	117.106	little	83	140	65.047	church
34	677	114.242	right	84	82	62.208	sister
35	293	113.733	give	85	356	61.19	better
36	351	112.952	thing	86	175	60.693	girl
37	340	112.627	help	87	392	59.226	put
38	265	112.2	feel	88	512	59.078	thought
39	107	111.852	lord	89	512	58.535	again
40	828	111.41	get	90	1183	58.292	now
41	113	109.594	playing	91	147	57.936	hold
42	805	108.609	know	92	1189	57.819	back
43	543	108.507	come	93	204	57.297	heard
44	205	101.11	actually	94	61	56.946	coach
45	807	100.39	good	95	263	56.61	job
46	248	99.681	won	96	79	56.426	pick
47	94	99.453	watch	97	157	55.567	hope
48	81	99.428	supposed	98	238	54.576	food
49	297	97.781	kind	99	165	54.547	remember
50	158	97.52	try	100	90	54.347	store

The show's focus on the everyday experiences of a young boy is reflected in the majority of keywords found, which are related to common daily activities and interactions with people around him. Keywords related to feelings or opinions are also dominant, indicating the show's focus on emotional expressions and interpersonal dynamics, and illustrating the characters' emotions and opinions. Other keywords found in the Young Sheldon series corpus are related to family, people, school, events and occurrences, highlighting the show's focus on familial relationships, educational settings, and significant events that influence the characters' lives and interactions. Additionally, there are a few keywords related to religion, which can be attributed to the influence of Sheldon's deeply religious mother. Therefore, the collection of the top 100 keywords can be organised into six main categories based on their meanings.

**Table 2***The Top 100 Keywords in the Young Sheldon Series Corpus*

<b>Meaning</b>	<b>Keywords</b>
Keywords related to family and people	brother, mother, sister, son, boy, girl, baby, kid, guy, young
Keywords related to school	school, football, science, coach
Keywords related to religion	lord, god, church
Keywords related to common daily activities	dinner, tell, talk, wait, think, eat, say, help, ask, need, playing, feel, watch, call, won, try, believe, understand, heard, stuff, go, here, look, give, thing, get, know, come, kind, problem, see, phone, read, night, take, sit, keep, door, stop, home, put, hold, job, pick, food, store
Keywords related to feelings or opinions	worry, love, sorry, nice, fine, fun, sure, wrong, cool, bad, little, good, great, interesting, better, really, pretty, want, sounds, right, supposed, kind, actually, mind, sweet, never, care, big, thought, hope, hell
Keywords related to events and occurrences	previously, happened, again, now, back, remember

The keywords were further classified into four grammatical categories based on their parts of speech as content words. These categories included nouns, verbs, adjectives, and adverbs, reflecting the diversity of their functions within the Young Sheldon series corpus.

**Table 3***The Top 100 Keywords in Four Grammatical Categories.*

<b>No</b>	<b>Function</b>	<b>Keywords</b>
1	nouns	stuff, boy, kid, dinner, baby, hell, thing, lord, kind, school, football, problem, brother, phone, mother, guy, night, son, science, door, home, god, church, sister, girl, coach, job, food



No	Function	Keywords
2	verbs	tell, talk, want, wait, go, love, sounds, worry, think, eat, say, look, ask, need, give, help, feel, get, playing, know, come, won, watch, supposed, try, believe, wants, gets, see, mind, understand, take, sit, keep, call, stop, care, happened, suppose, put, thought, hold, heard, pick, hope
3	adjectives	sorry, nice, fine, fun, sure, wrong, cool, bad, little, right, good, great, ready, young, sweet, big, interesting, better
4	adverbs	really, pretty, here, previously, actually, never, again, now, back

In Table 3, it is evident that in the TV series *Young Sheldon*, a significant portion of the keywords used are verbs, comprising 46 per cent of the total. Nouns follow closely behind, accounting for 27 per cent, while adjectives make up 18 per cent, and adverbs make up 9 per cent of the keywords used in the show.

#### 4.2 The Most Frequent Collocations of the Top 30 Keywords

After identifying the keywords in the corpus of the *Young Sheldon* series, the next step was to identify the lexical collocations. The top 30 keywords were used as “nodes” to identify their collocates using the collocate tool in AntConc. It generated a list of collocates within a four-word span to the left and right using log-likelihood ( $p < 0.05$ , 3.84 with Bonferroni) and Mutual Information ( $MI \geq 3$ ) measures. It selected only lexical collocations with a frequency of 10 or more. The results are outlined in Table 4.

**Table 4**

*Examples of the 30 Keywords with Their Collocations*

No	Keywords	Collocates	No	Keywords	Collocates
1	sorry	sorry (to) hear (that) sorry (I'm) late	16	dinner	having dinner eat (your) dinner
2	nice	nice (to) meet (you) nice try	17	go	get (to) go go home

No	Keywords	Collocates	No	Keywords	Collocates
3	fine	turned (out) fine	18	love	god (is) love (don't) believe (in) god
4	fun	have fun sounds fun	19	baby	having (a) baby baby shower
5	sure	make sure pretty sure	20	sounds	sounds like sounds good
6	tell	tell (you a) story tell (someone) the truth	21	hell	going (to) hell (What the) hell (were you) thinking?
7	stuff	(all) kinds (of) stuff science stuff	22	worry	(don't) have (to) worry (have) nothing (to) worry
8	talk	want (to) talk (about) need to talk	23	think	(Do you) ever think (of)/(about)...?
9	really	really want (Do you) really think...?	24	cool	pretty cool
10	wrong	something's wrong nothing wrong	25	eat	let's eat
11	boy	(my) little boy big boy	26	bad	feel bad bad news
12	kid	smart kid weird kid	27	say	say grace people say
13	want	want (to) hear want (to) play	28	look	look like look after
14	wait	wait (a) minute wait (in the) car	29	here	right here come here
15	pretty	pretty good pretty sure	30	ask	ask (you a) question

In Table 4, the list includes the first 30 keywords and their top two most frequent collocations. The pair will be selected based on a threshold of  $\geq 10$  occurrences and an MI value  $\geq 3$ . If no collocates met the criteria of having at least 10 occurrences and an MI value of 3 or higher, the criteria could be adjusted as necessary. While not all collocations occur directly next to the keywords, they do occur within a reasonable word span that is neither too close nor too far for each pair to co-occur.

This section categorized the top 30 keywords and their collocations from the “Young Sheldon” series based on the main meaning categories of the 100 keywords mentioned earlier: family and people, religion, common daily activities, and feelings or

opinions. This analysis highlighted the characters' interactions and important aspects of the show.

#### **4.2.1. Keywords and Their Collocations Related to Family and People**

**Boy:** The keyword “boy” is often used with the adjective “little” in reference to Sheldon. This term reflects the familial bond and parental care within Sheldon's family, particularly from his mother, Mary, who frequently refers to him as “my little boy.” The show emphasises his youth and the unique challenges of being a child prodigy. Occasionally, the show uses the term “big boy” to signify situations that reflect Sheldon's coming of age.

**Kid:** The word “kid” frequently appears with other adjectives. Phrases like “smart kid” and “weird kid” are used to emphasise Sheldon's unique qualities. These descriptions are central to the show's exploration of his character as a child prodigy who stands out due to his exceptional intellect and unconventional behaviour. They highlight the contrast between him and other children, revealing both his intelligence (‘smart kid’) and his social difficulties (‘weird kid’).

**Baby:** In the Young Sheldon series corpus, the phrases “having a baby” and “baby shower” often appear together as collocations. These phrases are important as they are related to the storyline about the birth of a baby, which is a major plotline in the series during seasons 5-6. Discussions about “having a baby” and events such as a “baby shower” are key moments that drive the narrative forward.

#### **4.2.2. Keywords and Their Collocations Related to Religion**

**Love:** The most frequent collocates that appear with the keyword “love” in the Young Sheldon series corpus are the noun “God.” and the adverb “much” as in “god (is) love” and “love (you so) much.” The phrase “God is love” is a recurring element in the show, reflecting religious themes through Mary Cooper's strong Christian beliefs. Moreover, the show emphasises the close and supportive relationships among the characters with the phrase “love you so much.”

**Say:** The most common collocations of the word “say” found in the Young Sheldon series corpus include “say grace” and “people say.” In the show, “say grace” reflects the Cooper family's religious practices, as they say grace before meals. This

phrase is frequently used in the dialogue. Another common collocation found in the script is “*people say*,” which is used to express common beliefs, opinions, or rumours. This reflects how the characters discuss social norms, gossip, and what people think, contributing to the show's social dynamics.

#### 4.2.3. *Keywords and Their Collocations Related to Common Daily Conversation*

**Tell:** The most common collocates of the keyword “*tell*” in the Young Sheldon Series Corpus are “*tell a story*” and “*tell (someone the) truth*.” These are frequently used when characters are sharing experiences or resolving conflicts. These combinations contribute to character development and provide background information. The show often focuses on conflicts that require communication, with characters needing to reveal the truth to clear up misunderstandings

**Stuff:** “*Young Sheldon*” uses casual and relatable conversations to reflect how people speak in everyday life, especially children and families. The term “*stuff*” is commonly used as a placeholder word in informal speech. It's often used when characters want to generalise or simplify their discussion. For example, phrases like “*(all) kinds (of) stuff*” or “*science stuff*” allow characters to refer to broad categories without going into specifics.

**Talk:** The keyword “*talk*” reflects conflict resolution, as many episodes in the Young Sheldon series involve resolving misunderstandings, conflicts, or emotional issues. Talk collocates with verbs on the left span like “*want (to) talk (about)*” and “*need (to) talk*”, are common when characters address problems, share feelings, or discuss important matters

**Want:** The most frequent collocates of the word “*want*” are “*hear*” and “*play*.” The characters in the show frequently express their desires and goals using phrases such as “*want to play*” and “*want to hear*.” These expressions, like “*want to hear*,” show curiosity or a desire for information, and “*want to play*,” indicates a desire for interaction or entertainment.

**Wait:** The word “*wait*” typically collocates with nouns such as “*wait (a) minute*” and “*wait (in a) car*.” In the show, characters use “*wait a minute*” to pause and reconsider a situation or statement. This phrase is used when someone needs a moment to process information, leading to a brief interruption or clarification in

conversation. Additionally, telling someone to “*wait in the car*” usually ensures safety and maintains control over a situation.

**Dinner:** “*Having*” and “*eating*” are collocations to the left of “*dinner*”, as seen in “*having dinner*” and “*eat (your) dinner.*” The collocations above describe the family-centric setting in the show “Young Sheldon,” which focuses on family dynamics. The dinner table is a frequently recurring setting in the show and serves as a common space for family gatherings and interactions. Phrases such as “*having dinner*” and “*eating your dinner*” highlight typical family activities and parental involvement, contributing to the authenticity of the show.

**Go:** The word “*go*” is essential for describing characters’ movements and actions. It frequently appears in different combinations in the Young Sheldon Series Corpus, such as directional collocations like “*go home*” or decision and permission collocations like “*get (to) go.*” These phrases are commonly used to show daily routines and activities.

**Eat:** The most frequently used collocation of the keyword “*eat*” in the Young Sheldon series corpus is “*Let’s eat.*” This is likely due to the importance of family meals to the Cooper family, which is often emphasised with phrases like “*let’s eat.*” It reflects the significance of shared mealtimes in their daily routine and relationships.”

**Look:** “*Look like*” and “*look after*” are the most common collocations of the word “*look*” in the Young Sheldon series corpus. These descriptive phrases describe appearances, make comparisons, and convey the context of caring for others. They are important for setting the scene and helping viewers visualise and understand the context, particularly in a family-centred show where characters often discuss responsibilities and the well-being of family members.

**Here:** The collocations “*right here*” and “*come here*” are frequently used in the Young Sheldon Series to establish location. The word “*here*” is often used to indicate where characters or objects are located. Phrases like “*right here*” help to specify particular places or situations, which is crucial for setting the scene and focusing the audience's attention. In addition, phrases such as “*come here*” are used to direct characters’ movements, particularly in family and school settings where parents or teachers need to gather or redirect children.

**Ask:** In the TV show *Young Sheldon*, the phrase “*ask a question*” is the most commonly used collocation with the word “*ask*.” This reflects the characters’ curiosity, especially that of Sheldon. As a highly intelligent child, Sheldon often asks questions to satisfy his curiosity and understand the world around him. What characters ask and the way they ask can also reveal a lot about their motivations, insecurities, and relationships with others.

#### 4.2.4. Keywords and Their Collocations Related to Feelings or Opinions

##### 4.2.4.1 Feelings

**Sorry:** The most frequent collocates within a four-word span to the right of the keyword *sorry* are “*hear*” and “*late*.” For example, “*sorry (to) hear (that)*” and “*sorry (I’m) late*.” These phrases commonly express politeness and consideration for others, reflecting good manners and respectful behaviour.

**Nice:** The word “*nice*” is often used alongside words like “*meet*” and “*try*,” as in “*nice to meet you*” and “*nice try*.” “*Nice try*” is often used sarcastically to indicate that the effort was inadequate or amusingly unsuccessful. On the other hand, “*nice to meet you*” is a straightforward, polite expression used when meeting someone for the first time, indicating a friendly and respectful greeting.

**Fine:** The word “*fine*” is commonly used in everyday conversations to indicate satisfactory or acceptable outcomes. In the context of the show “*Young Sheldon*,” characters often use “*fine*” to describe situations, events, or outcomes that meet expectations or turn out well. This is evident in collocations like “*turned (out) fine*”.

**Fun:** “*Have*” and “*sounds*” are the two most frequent collocations to the left of the keyword “*fun*,” as seen in pairs like “*have fun*” or “*sounds fun*.” The use of the word “*fun*” in the *Young Sheldon* Series corpus demonstrates how it is used in cultural and linguistic contexts. In “*Young Sheldon*,” characters frequently use “*fun*” to emphasise activities or situations that bring enjoyment, amusement, or leisure. Understanding the usage of “*fun*” in phrases like “*have fun*” or “*sounds fun*” can help language learners understand cultural norms and expressions in English-speaking contexts

**Sure:** The two most frequent collocations pairing to the left of the word “*sure*” are “*make sure*” and “*pretty sure.*” These combinations are frequently used in the show to describe character interactions. In “Young Sheldon,” characters often express certainty or doubt, resulting in the frequent use of phrases like “*make sure*” and “*pretty sure.*” These phrases are part of everyday language and help clearly convey the characters’ intentions and emotions. “*Make sure*” also reflects parental guidance, teacher instructions, and general advice, which are common in a family-centric show.

**Really:** The collocations “*really want*” and “*(do you) really think...?*” convey emphasis. “*Really want*” emphasises the intensity of a character’s desire or need. “*(Do you) really think...?*” expresses doubt or surprise. Phrases like this question the validity or likelihood of a statement, often indicating doubt or surprise. This helps characters convey their disbelief.

**Wrong:** The show often focuses on resolving conflicts, with characters frequently expressing that “*something’s wrong*” when identifying problems. This leads to dialogue aimed at resolving issues. Sheldon’s parents, teachers, and other authority figures often need to address problematic behaviours or situations. Therefore, they use collocations like “*nothing wrong*” to reassure him or “*something's wrong*” to point out issues that need to be corrected.

**Pretty:** It is a common colloquial intensifier in English and is used in everyday conversation as an adverb. It is often used to moderate adjectives, mostly collocating with adjectives as seen in “*pretty good.*” and “*pretty sure.*” Instead of describing something as excellent or certain, characters might say it’s “*pretty good*” or “*pretty sure.*” This reflects a balanced and nuanced way of communicating opinions and observations, indicating confidence without being absolute.

**Hell:** In the Young Sheldon series corpus, the word “*hell*” frequently appears alongside present participle verbs like “*going*” and “*thinking*”, as in the phrase “*going (to) hell*” and “*(What the) hell (were you) thinking?*” The phrase “*going to hell*” can convey a strong moral or religious judgment used by characters to emphasise the consequences of wrongdoings. “*Going to hell*” can also be strong language expressing frustration without implying a literal belief in hell. The phrase “*What the hell were you thinking?*” is often used in moments of tension or frustration. It expresses disbelief or exasperation over someone's actions.



**Worry:** The characters in the series frequently use reassuring phrases such as telling someone they “(don't) have (to) worry” or that they “(have) nothing (to) worry (about).” These phrases provide comfort and support, which is a recurring theme in the series due to Sheldon's peculiarities and the family's dynamics.

**Cool:** Characters often use the word “cool” to express approval or praise. In this corpus, the word “cool” is mostly paired with the adverb “pretty,” which collocates to the left of “cool.” This combination is used to emphasise the meaning of “cool”. Phrases like “pretty cool” are frequently used to compliment or recognise something impressive or praiseworthy, particularly in interactions among children, parents, and teachers.

**Bad:** In the Young Sheldon series, the most common collocations of the word “bad” are “feel bad” and “bad news.” These combinations suggest emotional expression, conflict, and drama. Characters often express their feelings using phrases like “feel bad” when discussing guilt, regret, sympathy, or disappointment. “Bad news” frequently introduces or highlights conflicts and challenges within a story.

#### 4.2.4.2 Opinions

**Sounds:** Collocates to the right of sounds suggest two kinds of meanings. The first one is conversational agreement and approval, as in “sounds good.” Characters often use this phrase to show agreement or approval of a suggestion or plan. The second meaning is to form impressions and judgments, as in the phrase “sounds like.” This phrase indicates a perception or impression based on what has been heard or understood.

**Think:** In the Young Sheldon series, the word “think” is often accompanied by the adverb “ever.” as in the phrase “(Do you) ever think (about/of) ...?” Characters frequently reflect on their experiences, beliefs, and feelings. Phrases like “Do you ever think about/of...?” encourage deep thinking. This indicates that characters in the show often engage in conversations to exchange thoughts and opinions.

### 4.3 Collocation Types and Combinations

Collocations give us insight into how words are commonly used together in natural language. It's important to identify and understand the types of collocations to



better understand the context in which these keywords appear. Understanding these relationships helps us comprehend how meanings are formed in dialogue and narrative and how words are related to each other. Hence, Table 5 contains examples of collocations, organised according to their combination types as suggested by Benson, Benson, and Ilson (2010).

**Table 5**

*Examples of Lexical Collocations Within Four-Word Span Based on Their Combination Types*

<b>Combinations</b>		<b>Examples</b>
Noun	Verb	people say
Verb	Noun	have fun, tell (you a) story, tell (someone the) truth, wait (a) minute, having dinner, eat (your) dinner, go home, (don't) believe (in) god, say grace
Adjective	Noun	(my) little boy, big boy, smart kid, weird kid, bad news
Noun	Noun	(all) kinds (of) stuff, science stuff, god (is) love, baby shower
Adverb	Adjective	pretty good, pretty sure, pretty cool
Verb	adjective	sounds fun, make sure, feel bad, sounds good, turned out fine
Adjective	Verb	sorry (to) hear (that), nice (to) meet (you)
Adverb	verb	really want, (Do you) really think...?, (Do you) ever think (of)/(about)...?
Verb	Verb	want (to) talk (about), need (to) talk, want (to) hear, want (to) play, let's eat
Adverb	Adverb	right here
Verb	Adverb	come here
Adjective	Adjective	sorry (I'm) late

The data in Table 4 indicates that the majority of the collocation pairs consist of Verb+Noun combinations (26%), followed by Verb+Verb (16%) and Adjective+Noun (14%) combinations. Both Noun+Noun and Adverb+Adjective, as well as Verb+Adjective combinations, account for 9% each. Additionally, Adverb+verb combinations make up 6%, and Adjective+Verb combinations make up 5%. The remaining combinations, which include Adverb+Adverb, Verb+Adverb, and Adjective+Adjective, each account for 2%.

This chapter illustrated and analysed the top 100 keywords and the collocations of the top 30 keywords found in the “Young Sheldon” series. The analysis highlighted how these keywords and their collocations relate to daily conversation, family, feelings and opinions, and religious themes, enhancing character and plot development. The findings support using media-based materials for effective language learning, providing practical examples of context-specific vocabulary and natural language use.



## **CHAPTER 5**

### **DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS**

This chapter is organised into five sections to provide a comprehensive overview of the study. These sections include (1) a summary of the study, (2) a summary of the findings, (3) pedagogical implications derived from the study, (4) limitations encountered during the study, and (5) recommendations for further research in the field.

#### **5.1 Summary of the Study**

This research investigated the use of keywords and their lexical collocations in the TV series “Young Sheldon” to improve English language learning and teaching. The study aimed to identify a list of 100 keywords from “Young Sheldon” transcripts and examine the most frequent collocations of the top 30 keywords. This study used AntConc software for corpus analysis, comparing the Young Sheldon Series Corpus with a reference corpus (AME06) to identify statistically significant keywords and their collocations.

The results revealed various keywords categorised into nouns, verbs, adjectives, and adverbs, representing different aspects of everyday life, school, religion, and common activities portrayed in the show. A detailed analysis of the collocations illustrated how these words are commonly used together in natural language to offer insights into their contextual meanings and relationships.

#### **5.2 Summary of the Findings**

This study is centred on identifying and interpreting keywords and their collocations in the TV series “Young Sheldon” for potential use as teaching material for vocabulary in context. Using AntConc software, the study carefully extracted and examined the top 100 keywords, focusing exclusively on content words. These keywords were categorised into four grammatical groups: nouns, verbs, adjectives, and adverbs. Additionally, the study identified the collocations of the top 30 keywords and provided interpretations of the contexts in which these collocation pairs commonly occur. In summary, the study's findings were outlined as follows:

### **5.2.1 The 100 Keywords**

From the 100 keywords extracted using AntConc software, each word was categorised into parts of speech categories: nouns, verbs, adjectives, and adverbs. This classification helped highlight the specific roles these keywords play in the dialogue of the TV series *Young Sheldon*, providing a clearer understanding of the linguistic structure and thematic emphasis in the show's scripts.

Verbs constituted a significant 46% of the keywords, indicating that the dialogues in the series are heavily focused on actions and activities. The characters in the series frequently talked about what they were doing, what they planned to do, or what others were doing. This action-oriented focus in their conversations made the dialogue dynamic and engaging, as it centred around performing and describing various tasks and activities. Examples include words like “tell,” “want,” “go,” and “say,” which highlight the dynamic interactions between characters.

Nouns constituted 27% of the keywords, highlighting important themes such as “stuff,” “school,” “mother,” and “science.” It was observed that over a quarter of the most frequently used words in the TV series “*Young Sheldon*” are nouns. This suggests that the dialogue in the series often revolves around specific subjects and important elements. Words like “stuff,” “school,” “mother,” and “science” are commonly used, emphasising their significance within the conversations. These nouns play a crucial role in establishing and reinforcing the main topics and key elements that the characters discuss. This helps to highlight the central themes and important aspects of their lives and interactions within the series.

A notable portion of the most frequently used words in the TV series “*Young Sheldon*” are adjectives and adverbs. Specifically, adjectives make up 18% of these keywords, and adverbs make up 9%. Adjectives are words that describe or modify nouns, providing details about the attributes, qualities, or characteristics of people, places, or things. Examples might include words like “nice,” “fine,” or “bad.” On the other hand, adverbs modify verbs, adjectives, or other adverbs, indicating how, when, where, or to what extent something happens. Examples might include words like “really,” “pretty,” or “here.” These adjectives and adverbs make the dialogue richer by describing different qualities and states, adding details about actions and situations.

The findings of the 100 keywords from the “Young Sheldon” series align closely with the previous studies discussed in Chapter 2, reinforcing the importance of authentic media in language learning. The high proportion of verbs in the keywords, reflecting an action-oriented nature, is consistent with Sezgin and Öztürk's (2020) study, which emphasised the dynamic use of language in TV series to enhance vocabulary and communication skills.

Similarly, the significant presence of nouns related to key themes such as family, school, and science mirrors Scheffler, Jones, and Domińska's (2021) findings in their analysis of “Peppa Pig,” which highlighted the use of thematic vocabulary relevant to the show’s context. In both studies, the frequent use of thematic vocabulary reflects the shows’ context, making them valuable for teaching relevant and context-specific language. For instance, “Young Sheldon” often revolves around family dynamics and scientific topics, similar to how “Peppa Pig” includes child-friendly topics like animals and weather.

Additionally, Patumanon’s (2016) study on intensifiers in “Friends” provides a relevant comparison for the use of intensifiers in “Young Sheldon.” Both studies reveal that intensifiers like “really” and “pretty” frequently emphasise and convey degrees of certainty or intensity in characters’ speech. This usage helps learners understand how to express intensity and nuance in everyday conversations.

### ***5.2.2 The Collocations of the Top 30 Keywords***

The collocations reveal patterns in how these keywords are used in natural language, providing insights into their contextual meanings and relationships. The top 30 keywords and their collocations from the “Young Sheldon” series were analysed into main meaning categories: family and people, religion, common daily activities, and feelings or opinions. Analysing these keywords and their collocations revealed insights into how language is used in the show to convey actions, emotions, and interactions.

The majority of the keywords fall into the categories of common daily activities and feelings or opinions. Keywords related to common daily activities, such as “tell a story,” “wait a minute,” and “eat dinner,” illustrate everyday conversations and routines, providing insights into the characters’ daily lives and interactions. Keywords

related to feelings and opinions, such as “really want” and “feel bad,” capture characters’ emotional expressions and interactions, revealing how they navigate their relationships and personal challenges.

Keywords related to family and people highlight familial bonds and Sheldon’s unique qualities as a child prodigy. Examples include “my little boy,” “big boy,” “smart kid,” and “weird kid,” which underscore the show’s focus on the challenges and dynamics within Sheldon’s family. Keywords related to religion, such as “God is love” and “say grace,” reflect the importance of faith and shared beliefs in the characters’ lives.

In *Young Sheldon*, collocations like “smart kid,” “weird kid,” and “my little boy” portray Sheldon’s intelligence, social challenges, and family dynamics. These phrases highlight his unique childhood prodigy experiences and his mother's caring role. Similarly, Siripachanah’s (2022) study on “Desperate Housewives” found that keywords related to family, relationships, and domestic roles play a significant role in portraying the protagonists’ personalities and the themes of domestic responsibilities and female unity. Just as the keywords and their collocations in “*Young Sheldon*” emphasise the familial and personal aspects of Sheldon’s life, the keywords in “*Desperate Housewives*” highlight the pressures on women’s domestic roles and the unity among female characters, showing how they navigate their complex lives.

The collocations of “love” with “God” and “say” with “grace” in “*Young Sheldon*” reflect the show’s religious themes. This is comparable to Hamed’s (2021) study on US presidential discourse, which showed how keywords and their collocations reveal underlying themes and priorities. In both studies, the collocations help convey the characters’ beliefs and the cultural context, emphasising the importance of understanding collocations for a deeper grasp of language use.

The frequent collocations in daily conversations in “*Young Sheldon*,” such as “having dinner,” and “want to play,” reflect everyday language use. Nation (2006) emphasised the need for learners to master common vocabulary for effective communication. The use of these terms in various contexts within the show mirrors Nation’s idea of essential vocabulary lists that cover a broad range of daily activities, reinforcing the practical application of commonly used words.

Overall, the analysis of the “Young Sheldon” series shows how valuable TV shows can be for learning languages. The show’s collocations give practical examples of how words are used together, helping learners improve their vocabulary, understand cultural context, and communicate better.

### **5.3 Implications**

This research explores the use of “Young Sheldon” TV series transcripts as effective resources for language teaching. Identifying the top 100 keywords and their collocations provides educators with practical vocabulary for lesson plans. The study emphasises the importance of authentic, context-rich content to enhance student engagement and language proficiency. The results of the study can also serve as a basis for designing interactive activities that reinforce vocabulary learning, making language use more natural and fluent. The following sections will detail these resources and suggest interactive activities to improve language learning outcomes.

#### ***5.3.1 Resources for Language Teaching***

The study’s identification of the top 100 keywords and their collocations provides teachers with the commonly used vocabulary found in the Young Sheldon series corpus that can be integrated into lesson plans, helping students acquire practical and frequently used language. Analysing these keywords and their collocations provides a good foundation for vocabulary instruction. The use of TV series transcripts immerses students in authentic language as it naturally occurs, aiding their understanding of how words and phrases are integrated into conversations.

Teachers can use “Young Sheldon” transcripts to expose students to authentic, contextual language, enhancing their comprehension by demonstrating how words and phrases are used in real-life situations. This is because the identified keywords represent everyday interactions, making them highly practical for learners and significantly enhancing their ability to communicate effectively in real-world situations.

Furthermore, the collocations of the top 30 keywords provide context-rich examples of word pairings, leading to more fluent and natural language use. Learning vocabulary through engaging and familiar content like a TV series improves memory,

as students are more likely to remember and use words encountered in meaningful and entertaining contexts.

Additionally, the study highlights vocabulary relevant to cultural themes and everyday life in English-speaking environments, helping students understand the context in which the language is used. Using a popular TV series as a learning tool can increase student engagement and motivation, supporting the development of speaking, listening, reading, and writing skills. Understanding and correctly using collocations can improve both fluency and accuracy, making students' language use sound more natural.

#### **5.4 Limitations of the Study**

The study focuses on a specific type of spoken English found in the TV series "Young Sheldon." The cultural and situational context in the show, set in East Texas, USA, reflects the Southern American English dialect, including distinctive regional expressions, idioms, and pronunciation patterns. This setting offers insights into regional language use and enriches learners' understanding of cultural and linguistic diversity. However, the findings may not be applicable to other contexts or media. Moreover, concentrating solely on keywords and their collocations may overlook other important language aspects, such as syntax, discourse features, and pragmatic nuances. Therefore, the study might provide a limited view of the language used in "Young Sheldon" by not addressing these broader linguistic features.

#### **5.5 Recommendations**

To build on the findings of this study, several recommendations are proposed for future research. These suggestions aim to enhance the understanding and application of media in language teaching, ensuring a more comprehensive and effective approach to language learning.

##### ***5.5.1 Broader Media Analysis***

Future studies should consider analysing a broader range of TV series, films, and other media to capture a wider variety of spoken English contexts and genres. This approach would help generalise the findings and provide a more comprehensive



understanding of language use. Including media from different regions and dialects of English can explore linguistic diversity, offering insights into regional variations and enhancing learners' appreciation of the language's varieties.

### ***5.5.2 Expanded Focus***

Expanding the focus beyond keywords and collocations to include syntax, discourse markers, and pragmatic elements would provide a more holistic view of language use. This broader focus would ensure a deeper understanding of how language operates in different contexts, enriching the analysis with insights into sentence structure, coherence, and contextual meanings.

### ***5.5.3 Incorporate Qualitative and Quantitative Approaches***

Additionally, incorporating qualitative or quantitative approaches, such as learner feedback, classroom observations, or longitudinal studies on vocabulary memory, could offer deeper insights. Studies comparing the effectiveness of TV series with other types of media, such as podcasts, news broadcasts, and social media, can help identify the best tools for language learning. Furthermore, research could investigate how TV series can be integrated with other educational methods, such as interactive language apps, classroom activities, and online platforms, to create a comprehensive learning experience. Addressing these areas in future research can build on the findings of this study, further enhancing our understanding of how media can be used effectively to teach and learn languages.

## REFERENCES

- Anthony, L. (2005, July). *AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom*. In IPCC 2005. Proceedings. International Professional Communication Conference, 2005. (pp. 729-737). IEEE.
- Baker, P. (2007). *AmE06 in AntConc format*. The AmE06 Corpus. <https://www.lancaster.ac.uk/linguistics/about/people/paul-baker>.
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh University Press.
- Biber, D., & Reppen, R. (Eds.). (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press.
- Cheng, W. (2013). Semantic prosody. *The encyclopedia of applied linguistics*, 1-7.
- Conzett, J., Hargreaves, P. H., Hill, J., Lewis, M., & Woolard, G. C. (2000). *Teaching collocation: Further developments in the lexical approach* (pp. 163-199). M. Lewis (Ed.). Boston: Thomson Heinle.
- Egbert, J., & Biber, D. (2019). Incorporating text dispersion into keyword analyses. *Corpora*, 14(1), 77-104.
- Firth, J. R. (1935). The technique of semantics. *Transactions of the philological society*, 34(1), 36-73.
- Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in English*. Routledge.
- Hamed, D. (2021). Keywords and collocations in US presidential discourse since 1993: A corpus-assisted analysis. *Journal of Humanities and Applied Social Sciences*, 3(2), 137-158.
- Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge University Press.
- Kyröläinen, A. J., & Laippala, V. (2023). Predictive keywords: Using machine learning to explain document characteristics. *Frontiers in Artificial Intelligence*, 5, 975729.

- Lengerlioglu, Y. (2019). *The effects of TV series as curricular activities on speaking skills of English language learners* [Master's thesis, Eğitim Bilimleri Enstitüsü].
- McEnery, T., & Wilson, A. (1997). Teaching and language corpora (TALC). *ReCALL*, 9(1), 5-14.
- Miščin, E., & Miščin, S. (2018). Using Tv shows in teaching English. *Strani jezici: časopis za primijenjenu lingvistiku*, 47(1-2), 103-110.
- Nation P. (2013) Materials for teaching vocabulary. In: Tomlinson, B (ed.) *Developing materials for languages teaching*. (pp.351-365). Bloomsbury.
- Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, 63(1), 59-82.
- O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Patumanon, M. W., & Singhapreecha, P. (2016). *A comparative corpus study on intensifiers in American spoken language and tv dialogues based on the tv series "Friends"* [Doctoral Dissertation, Thammasat University].
- Rayson, P., & Chapelle, C. (2019). Corpus analysis of key words. *The Concise Encyclopaedia of Applied Linguistics*, 320-326.
- Riemer, N. (2010). *Introducing semantics*. Cambridge University Press.
- Scheffler, P., Jones, C., & Domińska, A. (2021). The Peppa Pig television series as input in pre-primary EFL instruction: A corpus-based study. *International Journal of Applied Linguistics*, 31(1), 3-17.
- Sezgin, H., & Öztürk, M. S. (2020). A corpus analysis on the language on TV series. *Journal of Language and Linguistic Studies*, 16(1), 238-252.
- Siripachanah, U. (2022). *A corpus stylistic study of the TV series script desperate housewives* [Master's independent study, Chulalongkorn University].
- Stewart, D. (2010). *Semantic prosody: A critical evaluation*. (Vol. 9). Routledge.

- Szczygłowska, T. (2022). Exploring Obama's and Trump's political discourse through the lens of wordlists, keywords and clusters. *Brno Studies in English*, 48(1).
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge.
- Tutin, A. (2008, July). For an extended definition of lexical collocations. In *Euralex* (p. 00).
- Wahyuni, R., & Oktaviani, O. (2021). Lexical and grammatical cohesion in The Lion King movie script. *English Teaching and Linguistics Journal (ETLiJ)*, 2(1), 133-142.
- Williams, R. (2014). *Keywords: A vocabulary of culture and society*. Oxford University Press.
- Zethsen, K. K. (2006). Semantic prosody: Creating awareness about a versatile tool. *Tidsskrift for Sprogforskning*, 4(1), 275-294.