

---

| RESEARCH ARTICLE

## Are TED Talks Suitable Materials for Humanities Specialized Vocabulary Learning?

Xiaotong Zhang

BA student in English Department of Applied English, School of Foreign Languages and Literature, Shandong University, Jinan, China

**Corresponding Author:** Xiaotong Zhang, **E-mail:** [xiaotongzhang@mail.sdu.edu.cn](mailto:xiaotongzhang@mail.sdu.edu.cn)

---

| ABSTRACT

Previous studies have revealed the pedagogical significance of TED talks for academic vocabulary learning. However, their suitability as learning materials for humanities students remains unclear. This study will use a corpus-based method to investigate whether TED talks are appropriate materials for humanities students to learn specialized spoken vocabulary. A Humanities TED talks Corpus (HTTC) and five sub-corpora of different subjects are developed as research objects. The coverage of three wordlists over the HTTC and sub-corpora is analysed and compared, including Academic Word List (AWL), Academic Spoken Word List (ASWL), and Soft Science Word List (SSWL). Levels 2-4 coverage of ASWL and SSWL is analysed with emphasis because they can better represent academic lexis. An additional corpus of TED Ed is also developed to advance the current investigation. The findings revealed that SSWL had higher coverage (86.2% overall, 8.9% in levels 2-4) than ASWL (85.2% overall, 7.7% in levels 2-4), which suggested that humanities related TED talks can be potential materials for specialized spoken vocabulary learning. The coverage comparison among sub-corpora reported the highest coverage of History, which was discovered due to the large percentage of TED Ed talks, presumably a more technical speech event of TED talks. This study may provide pedagogical instruction for introducing TED talks as supplementary materials into English for specific academic purpose (ESAP) classes to benefit humanities students' specialized spoken vocabulary learning.

| KEYWORDS

TED talks; lexical coverage; corpus; humanities

| ARTICLE INFORMATION

**ACCEPTED:** 01 October 2022

**PUBLISHED:** 06 October 2022

**DOI:** 10.32996/ijllt.2022.5.10.6

---

### 1. Introduction

TED talk is a database of speeches that are largely delivered by renowned specialists from various fields, including scientists, enterprisers, writers, statesmen, artists, and a great number of scholars (Cassidy et al., 2013). Those speakers are regarded as thinkers, doers, and idea-creators who devote themselves to 'make great ideas accessible and spark conversation' through short powerful talks. Besides, a wide diversity of topics across different disciplines are publicly available on TED official website, such as Technology, Science, Global Issues, Culture, Design, etc. Due to these features, it is common to employ TED talks in educational practices, especially in English programs in universities.

Previous studies have established the pedagogical significance of TED talks for students' English learning, like upgrading their speaking skills (Nurmukhamedov, 2017) and enhancing their academic listening abilities (De Chazal, 2014). Moreover, based on the analysis of lexical coverage, a number of studies have found the potential of TED talks for students' spoken academic vocabulary learning in English for Academic Purpose (EAP) classes, and it has been verified that this potential can vary according to talks of different disciplines (Liu & Chen, 2019; Wingrove, 2021). Considering that different disciplines have different needs for academic vocabulary and TED talks provide a wide variety of interdisciplinary topics, they may not be appropriate materials for all students to expand academic vocabulary. For example, previous studies have pointed out that the vocabulary needs for humanities can be very different from other disciplines (Durrant, 2009), and whether humanities related TED talks are less suitable for academic vocabulary learning should be further explored (Wingrove, 2021). Hence, investigating the potential of TED talks can be conducive

to humanities students' academic vocabulary learning, the avoidance of misleading teaching, and exploring new pedagogical knowledge of TED talks.

Despite this significance, few studies have investigated humanities related TED talks so far, and they largely concentrated on TED talks for general academic vocabulary learning rather than specialized vocabulary learning. Therefore, to examine the suitability of TED talks for humanities specialized vocabulary learning, the current study analysed the lexical coverage of three wordlists with a corpus-based method. The Humanities TED Talks Corpus (HTTC) and five sub-corpora of different subjects were developed, and three lists, Academic Word List (AWL), Academic Spoken Word List (ASWL), and Soft Science Word List (SSWL) were applied to calculate the coverage.

## **2. Literature Review**

In this section, the author first introduced lexical coverage and its relationship with the comprehension of academic speech. Then, previous studies on academic wordlists and the lexical profile of TED talks were reviewed and discussed. Finally, the research aims and questions of the current study were proposed by the author.

### **2.1 Lexical coverage and the comprehension of academic speech**

Lexical coverage, which is defined as "the percentage of running words in the text known by the readers" by Nation (2006, p. 61), is a useful measure to measure the vocabulary size needed for comprehending written or spoken discourse. Generally, 95% coverage is necessary for students to reach an adequate comprehension and guess unknown vocabulary in texts, and 98% coverage may be needed for optimal listening comprehension (Liu, & Nation, 1985; Stæhr, 2009; van Zeeland & Schmitt, 2013). Moreover, these figures are acceptably used in the coverage study of academic spoken discourse. For example, Dang and Webb (2017) chose 95% and 98% to indicate a lower and higher-level understanding of academic spoken English. And they were also used by Liu to measure the vocabulary needs for medicine-related TED talks (Liu, 2022).

### **2.2 From general wordlists to specific wordlists**

Considering the close connection between lexical knowledge and reading and listening comprehension, researchers have developed several academic lexis lists to help improve students' comprehension of academic discourse. For example, Academic Word List (AWL) and Academic Vocabulary List (AVL) were developed by Coxhead (2000) and Gardner and Davies (2014) respectively, which both focus on academic lexis in writing discourse. Moreover, Dang (2017) developed the first Academic Spoken Word List (ASWL) to help L2 learners expand their spoken vocabulary, irrespective of the differences in disciplines, vocabulary proficiency, or department structures. Based on these lists, a number of studies examined their validity across disciplines and discourse types. Take ASWL for instance, it has an average coverage of around 90% across hard and soft disciplines (Dang, 2017) and similar figures were found in the OpenCourseWare (OCW) and Massive Open Online Courses (MOOC) corpora (Liu, 2021).

Although previous studies have verified the existence of a core vocabulary word list, many researchers have revealed the disadvantages of general word lists and suggested the development of discipline-specific lists. Hyland and Tse (2007) argued that although the AWL has about 10% coverage in their research, the meaning, frequency, distribution, and collocation of many vocabularies can vary greatly across disciplines. Dang (2018a, 2018b) also pointed out that high-frequency and wide-range lexis may be different from hard sciences to soft sciences, and discipline-specific wordlists may be more useful to ESAP students. As a result, a number of academic writing wordlists for specific English have been created based on general wordlists, like Chemistry Academic Word List (CAWL) (Valipouri & Nassaji, 2013), Medical Academic Vocabulary List (MAVL) (Lei & Liu, 2016), Economics Academic Word List (EAWL) (O'Flynn, 2020), etc. However, few such wordlists have been explored in the field of academic spoken discourse, except Dang's Hard Science Word List (HSWL) (2018a), Soft Science Word List (SSWL) (2018b), and Medical Spoken Word List (MSWL) (2019), and their legitimacy is needed to be further verified. This may result from the inconvenience of collecting and calculating spoken statistics (Adolphs & Schmitt, 2003; Thompson, 2006).

### **2.3 Lexical profile of TED talks**

TED talks are a series of speeches and presentations delivered by experts from various fields like business, education, sciences, and global issues. Considering the similarity between TED talks and academic discourse (Cassidy et al., 2013), they are widely introduced to the academic listening practice, particularly for students who are engaged in EAP classes (De Chazal, 2014). Nevertheless, whether TED talks are suitable supplementary resources for EAP classes is still debatable.

Hitherto, several studies have investigated the lexical coverage of TED talks to figure out this issue. For instance, the ASWL was used by Liu and Chen (2019) to examine the lexical profile of a TED corpus including 2085 speeches from six themes, which reported a similar ASWL coverage (90%) over TED corpora and academic speech, indicating the suitability of introducing TED talks into EAP courses.

Taking this step further, Wingrover (2021) compared the academic lexical coverage of three wordlists (AWL, AVL, ASWL) in TED Talk Corpora (TTC) and Open Yale Course Lecture Corpora (OYCLC) both overall and across disciplines. He found that although TED talks have lower coverage than academic speeches in most cases, their difference is slight concerning the rate of encounter and common academic lexis. So, he suggested that it is pedagogically potential to adopt TED talks as listening materials in English for general academic purposes (EGAP) classes. Despite this, Wingrover (2021) also noted that humanities have the lowest lexical coverage in both corpora, probably because of its unique vocabulary compared to other disciplines (Durrant, 2009). So, he pointed out that humanities related TED talks are less likely to benefit EGAP students.

Unlike previous studies that all concentrate on the suitability of TED talks for general academic vocabulary learning, Liu (2022) further investigated it regarding specialized vocabulary. He analysed the coverage of MSWL and the encounter times of these words in medicine related talks. It was reported that 96.2% of specialized words appear in TED talks approximately, and they were likely to be frequently encountered by students, indicating that TED talks can potentially benefit specialized vocabulary learning.

Although previous studies have carefully examined the suitability of introducing TED talks into EAP classes from a vocabulary aspect, certain limitations remain. First, Liu and Chen (2019) used ASWL (a general wordlist) to investigate the vocabulary across different topical categories. Given that each discipline has its technical lexis and some academic words may have a unique meaning in specific subjects, it may not be appropriate to examine the lexical profile of TED talks with general wordlists. At present, only Liu's (2022) study used a discipline-specific wordlist. Second, Wingrover (2021) pointed out that TED talks are less likely to help humanities students expand the academic lexis they need. However, he did not use a humanities-specific wordlist to examine the coverage concerning humanities, which may lead to a debatable conclusion. Thus, a specifically spoken wordlist is necessary to be used to investigate this issue in further research. Third, previous studies all focused on the potential value of TED talks for EGAP courses except Liu's (2022) study focusing on English for Specific Purpose (ESP), suggesting that TED talks are likely to benefit specialized vocabulary learning in the medicine field. However, whether TED talks are potential learning materials for students majoring in other specific fields, like humanities, is still unknown.

Taken together, no study has investigated the lexical coverage of humanities related TED talks with a discipline-specific spoken wordlist. And whether humanities related TED talks are suitable learning materials for Humanities students is still debatable. To fill the gap, this study will examine the coverage of AWL, ASWL, and SSWL in humanities related TED talks, and investigate the suitability of TED talks for humanities students to learn specialized spoken vocabulary.

Founded on the literature and research aim, three research questions were designed to conduct the current study:

- i. What is the lexical coverage of AWL, ASWL, and SSWL in humanities related TED talks?
- ii. Is there any difference in lexical coverage among different subjects of humanities?
- iii. To what extent are TED talks suitable materials for humanities students to learn specialized spoken vocabulary?

### 3. Method

#### 3.1 The Corpus

TED is a collection of speeches mostly given by experts from various fields. As stated by the TED website, its mission is 'to make great ideas accessible and spark conversation' in the form of short powerful talks. Talks since 2006 are freely available on ted.com which are mostly presented with videos and full transcripts provided by the official producers. TED Ed is a project launched by TED talks for education purposes, making "short video lessons worth sharing, aimed at educators and students". As a result, the language style of TED Ed can be supposed different from live speeches, which may be more formal and less colloquial. Most talks on the website are categorized by topics like technology, entertainment, and design, or speech events like TED conferences, TEDGlobal, and TED Ed. Each talk has several topic tags to show its disciplinary field, speech event, and main contents. For example, the talk "*Can stereotypes ever be good*" has five tags including TED Ed, Education, Animation, Race, and History, which means it is a history related TED Ed talk concerning the topic of race.

Considering that the current study will investigate to what extent can TED talks help humanities students learn specialized spoken vocabulary, humanities related TED talks will be selected as research materials. According to Britannica Academic, the humanities include the study of all languages and literature, history, philosophy, and the arts, which are distinguished from natural sciences and social sciences. Based on this definition and the existing topics provided by the TED website, talks with the tag of language, literature, art, history, and philosophy will be chosen as targets. Some TED talks have more than one tag mentioned above. For instance, "*Why is the Mona Lisa so famous*" can be targeted by both art and history. These kinds of talks were not included in the corpora, because the current study was conducted from a disciplinary perspective, the involvement of talks with multi tags may affect the comparison between subjects. To make the size of each sub-corpus appropriate and comparable, qualified talks from 2017 to 2021 are collected.

As shown in Table 1, a Humanities TED talks corpus (HTTC) consisting of five sub-corpora of different subjects was developed. It should be noted that the sizes of five sub-corpora were not well-balanced due to the unbalanced number of talks on five topics and the unbalanced tokens in each talk. The number of TED Ed was also presented in the table, which was useful for the analysis of coverage of specific subjects in the following section.

**Table 1** The HTTC and five sub-corpora

	Total number	TED Ed number	Tokens
Art	53	8	78,513
History	53	40	49,124
Language	37	7	51,465
Literature	24	14	37,311
Philosophy	48	14	95,003
HTTC	215	83	311,416

### **3.2 Data Analysis**

Three wordlists were used to analyse the lexical coverage of humanities related TED talks in the current study, including the Academic Word List (AWL), the Academic Spoken Word List (ASWL), and the Soft Science Word List (SSWL).

The AWL was developed by Coxhead from 3.5-million-word academic corpora with the criteria of specialized occurrence, range, and frequency. The selected lexis was organized in the form of 570 headwords or word families. Based on this list, a number of studies have verified the vocabulary coverage of AWL in different disciplines and text discourses, and several discipline-specific wordlists have been developed.

The ASWL was generated from a 13-million-word corpus consisting of various academic spoken discourses, including those from diverse disciplines, speech events, and varieties of English. The list contains 1,741 word families and they were further divided into four levels according to the frequency of headwords. The size of each level is shown in the table. Previous studies have verified the high academic feature and spoken features of ASWL. Its legitimacy of being used in this study can be ensured.

The SSWL, containing 1,964 word families, was developed from a 6.5-million-word corpus of academic speeches of soft sciences. Similar to ASWL, the word families in SSWL are also divided into four levels based on their frequency. Compared to ASWL, this list can better represent the lexis in soft sciences related academic speeches, which was established in Dang's (2018b) study that SSWL has a higher coverage in soft sciences speeches than ASWL. Thus, the SSWL may be a more appropriate list for analysing the specialized vocabulary in humanities related TED talks.

The results comparison between the ASWL and the SSWL was made to see whether humanities related TED talks contain more specialized vocabulary. The two lists were comparable for certain reasons. First, they were both comprised of flemmas, so the unit of counting is the same. Second, they were developed with similar methods around the same time and were divided into four levels under the same standard. Third, they both focus on academic spoken discourse, with discipline as the only variation. Flemma was taken as the unit of counting in the current study to analyse the coverage of two lists and the coverage of each level in the HTTC.

Besides, AntWordProfiler (version 2.0.0) was employed to run the ASWL and the SSWL over the HTTC due to its convenient functions. It is a freeware vocabulary profiling and editing tool allowing users to create their wordlist to analyse the lexical coverage of targeted files. It also allows users to view the coverage at each level of the list. This tool was thus used by the researcher to investigate the vocabulary profiles of HTTC and the five sub-corpora of different subjects.

## **4. Results and Discussion**

### **4.1 Overviews of Lexical Coverage in the HTTC Corpus**

This section aims to answer RQ1: What are the lexical coverage of AWL, ASWL, and SSWL in TED talks in terms of Humanity disciplines? Table 2 shows the coverage of AWL in the HTTC both overall and in five sub-corpora of different subjects respectively. As can be seen, the coverage of AWL across the HTTC is 4.0% in total, which was similar to the coverage (3.9%) reported by Coxhead and Walls (2012) over their TED talks corpus. This suggested that the current study can further verify the lexical profile of AWL in TED talks established by the previous study. Moreover, Dang and Webb (2014) also reported that the AWL had coverage of 3.82% over arts and humanities related texts of the British Academic Spoken English (BASE) corpus, which was slightly lower than its coverage in HTTC. This difference may reveal that TED talks cover more academic vocabulary than BASE corpus in terms of humanities disciplines.

**Table 2** Lexical Coverage (percentage) of AWL in the HTTC

List	Art	History	Language	Literature	Philosophy	HTTC
AWL	3.7	5.0	3.9	2.9	4.0	4.0

In terms of the coverage of AWL in five sub-corpora, it can be found that the AWL coverage in Literature and Philosophy is 2.9% and 4.0% respectively, which were mostly consistent with the coverage reported by Wingrove (2021) (3.28% and 3.36%). However, the coverage of History (5.0%) was evidently higher than that reported by the previous study (3.9%) (Wingrove, 2021). One reason for the discrepancy may be that the transcripts of TED Ed talks constituted most of the texts in the historical corpus, with 41 texts among 53 in total. As explained in the method section, the language style of TED Ed talks may be more formal than common speeches and hence likely led to the higher coverage of AWL over history in the current study.

**Table 3** Lexical Coverage (percentage) of ASWL in the HTTC

Level	Art	History	Language	Literature	Philosophy	HTTC
ASWL Level 1	77.8	70.8	78.6	74.8	79.3	77.5
ASWL Level 2	5.0	6.3	5.6	4.0	5.0	5.2
ASWL Level 3	2.2	3.0	2.5	1.6	2.5	2.4
ASWL Level 4	0.1	0.1	0.1	0.1	0.2	0.1
ASWL Total	85.1	80.2	86.8	80.5	87.0	85.2
ASWL 2-4	7.3	9.4	8.2	5.7	7.7	7.7

Table 3 shows the lexical coverage of ASWL in the HTTC and five sub-corpora as well. Overall, the ASWL coverage is 85.2% in the HTTC, with levels 1, 2, 3 and 4 covering 77.5%, 5.2%, 2.4% and 0.1% respectively. It was found that the total coverage (85.2%) in the current study was statistically lower than those reported in Wingrove (2021)'s study (89.6%). One possible explanation for its occurrence is that the disciplines of the current corpus were restricted to humanities, while the topics of Wingrove's corpus covered various disciplinary groupings, like science, tech, culture, design, etc. This finding indicated that the lexical coverage of ASWL can vary with disciplines, and humanities related TED talks may contain fewer academic spoken vocabulary than other topics.

According to Wingrove (2021), ASWL level 1 should be removed when considering to comprise a set of academic lexis because it contains a large amount of high-frequency general English. The current study has adopted a similar method to Wingrove's with which the researcher particularly measured the coverage of levels 2-4 to indicate the academic attribute of the vocabulary appearing in humanities related TED talks.

When compared to Wingrove's study in terms of the ASWL levels 2-4 coverage over humanities TED talks, it was found that the HTTC (7.7%), literature (5.7%), and philosophy (7.7%) showed similar coverage with the figures reported by Wingrove (2021), which was 7.16%, 5.32%, and 7.32% respectively. However, it should be noted that history showed evidently higher levels 2-4 coverage than that reported in the previous study (Wingrove, 2021). This can be explained by the fact that the current history corpus included a large proportion of TED Ed, a more formal sort of talk, making it contain more academic lexis and thus show a higher coverage. Moreover, the higher coverage of AWL in the history corpus mentioned above can also provide a possible explanation for this difference.

**Table 4** Lexical Coverage (percentage) of SSWL in the HTTC

Level	Art	History	Language	Literature	Philosophy	HTTC
SSWL Level 1	78.8	71.5	78.9	79.7	79.6	78.0
SSWL Level 2	5.5	7.2	6.1	4.5	4.3	5.7
SSWL Level 3	2.7	4.0	3.1	2.4	3.1	3.1
SSWL Level 4	0.1	0.1	0.1	0.1	0.2	0.1
SSWL Total	87.1	82.8	88.2	86.7	87.2	86.9
SSWL 2-4	8.3	11.3	9.3	7.0	7.6	8.9

As presented in table 4, the SSWL covered 86.9% vocabulary of HTTC, with a separated coverage of 78.0%, 5.7%, 3.1%, and 0.1% in levels 1, 2, 3, and 4, which appeared to increase than the figures measured by the ASWL. In terms of levels 2-4, the SSWL also showed a higher coverage than ASWL in the current study, with an evident increase of 1.2%, which indicated that the SSWL may cover more academic lexis of humanities related TED talks than ASWL, and it contributed to a number of specialized vocabulary

to these talks. Through this finding, we can infer that the SSWL can be a more suitable wordlist than ASWL for the instruction of using TED talks to learn humanities specialized vocabulary.

#### **4.2 Lexical Coverage of Specific Subjects**

In answer to the RQ2: Is there any difference in lexical coverage among different subjects of humanities, the author analyses the lexical coverage of five subjects-specific corpora in this section. As can be seen from table 3, the ASWL has the highest coverage in History and the lowest coverage in Literature. Tables 4 and 5 show that History measured the lowest coverage with both ASWL and SSWL, which were 80.2% and 82.8% followed by Literature (80.5% and 86.7%), while Philosophy and Language were reported to have the highest coverage of ASWL and SSWL respectively. However, it should be noted that the higher total coverage of ASWL and SSWL in certain subjects may not indicate that they contain more academic lexis or specialized vocabulary than those that have lower total coverage. To investigate this issue, further comparison between the lexical coverage of ASWL 2-4 and SSWL 2-4 is necessary.

As illustrated in figure 1, Literature has the lowest coverage of ASWL 2-4 and SSWL 2-4, which is consistent with its low total coverage mentioned above. This may be due to the reason that the vocabulary needed for literature TED talks is more specialized than in other humanities talks. By contrast, History reported the highest levels of 2-4 coverage, though its total coverages of two lists were the lowest among five sub-corpora. This inconsistency between levels 2-4 coverage and total coverage suggested a larger proportion of levels 2-4 lexis in the history corpus compared to other sub-corpora, which indicated that the vocabulary of history TED talks can be considered more academic than other humanities related talks because levels 2-4 coverage can better represent the distribution of academic lexis than including the coverage of level 1 (Wingrove, 2021). Moreover, this can further suggest the possible inconsistency of total coverage and levels 2-4 coverage in academic spoken discourse, also the latent inaccuracy of measuring academic lexis with total coverage.



**Figure 1** The Coverage of Levels 2-4 in Five Subject-specific Corporuses

One credible explanation for this inconsistent coverage is the large proportion of TED Ed in the history corpus. As explained above, the language of TED Ed can be considered more formal, containing more academic lexis than colloquial speeches. And it is possible that the transcripts of TED Ed are written before the presentation, adding the feature of academic written language which may make the TED Ed talks more academic. As a result, the large proportion of TED Ed made the history corpus seem more academic, with the lowest level 1 and total coverage, but the highest levels 2-4 coverage.

As shown in figure 1, Art and Language have relatively higher levels of 2-4 coverage than Literature and Philosophy, which should not be due to the influence of TED Ed, because there are fewer TED Ed talks in Art and Language than Literature and Philosophy. According to Wingrove's (2021) study, TED talks of natural sciences (7.42%) and social sciences (7.83%) appeared to have higher ASWL 2-4 coverage than humanities (7.16%), with certain disciplines showing evidently high coverage, like physics TED talks (8.11%) and economics TED talks (8.66%). Through this, we can infer that the higher levels 2-4 coverage of Art and Language may be attributed to their higher level of combination with natural sciences and social sciences. This can also be seen from the talks collected in the art corpus and language corpus which covered a wide range of disciplines and topics, while talks in the literature corpus and philosophy corpus showed a lower level of combination with other areas. This may indicate that the vocabulary of art

and language can be more easily linked with other disciplines, while the vocabulary needs for literature and philosophy may be more unique and exclusive.

From the general comparison of ASWL coverage and SSWL coverage in figure 1, it can be found that levels 2-4 coverage from ASWL to SSWL shows a varying increase over all sub-corpora, except philosophy showing a nearly equal levels 2-4 coverage between two lists. This may suggest that the vocabulary needs for philosophy TED talks were less specialized than in other humanities related talks. Besides, the SSWL 2-4 coverage in the history corpus shows an increase of 1.9% compared to ASWL 2-4 coverage, which was the highest among the five sub-corpora. This suggested that history TED talks of the current study may contain a more specialized vocabulary of humanities, which was consistent with the finding reported above that history TED talks contained more academic lexis. The reason for this may be also due to the large proportion of TED Ed, which was further testified in the following part. Generally speaking, the increase of levels 2-4 coverage from ASWL and SSWL further indicated that the SSWL can be a better instructional wordlist than ASWL for learning humanities specialized vocabulary through TED talks.

### 4.3 TED Talks for Specialized Vocabulary Learning

This section aims to answer RQ3: To what extent are TED talks suitable materials for humanities students to learn specialized spoken vocabulary? As analysed above, the AWL coverage in the history corpus was evidently higher than in other sub-corpora. Although History had the lowest level 1 coverage of ASWL and SSWL, its levels 2-4 coverages of two lists were the highest among five sub-corpora, which were clearly higher than the figures reported in the previous study as well. The reason for these unusual coverages has been attributed to the possible influence of the large proportion of TED Ed talks in the history corpus.

To investigate the legitimacy of this hypothesis, the researcher developed a History Ed corpus composed of 40 TED Ed talks from the original history corpus and examined its lexical coverage of AWL, level 1, and levels 2-4 of ASWL and SSWL respectively.

**Table 5** Coverage of History and History Ed

List and Level	History	History Ed
AWL	5.0	5.8
ASWL level 1	70.8	65.4
ASWL levels 2-4	9.4	10.8
SSWL level 1	71.5	66.1
SSWL levels 2-4	11.3	13.2



**Figure 2** Coverage of levels 2-4 over History and History Ed

Table 5 shows the coverage comparison between History and History Ed. As we can see, History Ed suggests an obviously higher AWL coverage and levels 2-4 coverages than History, while its level 1 coverages of two lists are significantly lower than History. These results were consistent with the previous findings that History had the highest AWL coverage, highest levels 2-4 coverage, and the lowest level 1 coverage among five sub-corpora. Hence, we can infer that the previous findings were due to the influence of TED Ed coverage. Moreover, as presented in figure 2, History Ed shows more growth of levels 2-4 coverage from ASWL to SSWL

than History, which is a 2.4% increase compared to a 1.9% increase in History. This suggested that the highest coverage increase of levels 2-4 in History possibly resulted in the high coverage increase of History Ed.

Accordingly, the previous hypothesis of the unusual coverages of History can be largely verified by this further investigation. It also provided an implication that history TED Ed can contain more academic and specialized vocabulary than common history TED talks which can be suitable materials for humanities students to learn specialized vocabulary in terms of spoken discourse.

## **5. Conclusion**

This study analysed the lexical coverage of AWL, ASWL, and SSWL in humanities related TED talks to investigate the suitability of TED talks for humanities students to learn specialized spoken vocabulary.

In terms of the overall coverage in the HTTC, the total coverage and levels 2-4 coverage of SSWL was 86.2% and 8.9% respectively, while the counterpart of ASWL was 85.2% and 7.7%, which was lower than SSWL. Moreover, levels 2-4 coverage from ASWL to SSWL also showed an evident increase over most of the sub-corpora. These findings indicated two conclusions that can contribute to pedagogical implications. First, the SSWL can be a more suitable wordlist than ASWL for the instruction of using TED talks to learn humanities specialized vocabulary. Second, humanities related TED talks may contain a large number of specialized academic lexis, and thus can be suitable learning materials for students in ESAP classes. The second conclusion can also promote Wingrove's (2021) finding that although TED talks may not be ideal materials for EGAP students, they can be possibly introduced to ESAP classes, bringing new pedagogical implications for EAP courses.

In terms of specific subjects, the relatively higher levels 2-4 coverage of Art and Language showed that the vocabulary of art and language can be more easily linked with other disciplines. Besides, the abnormally high levels of 2-4 coverage of history led to a further investigation on the coverage of History TED Ed talks in the current study, which showed that History Ed had higher levels of 2-4 coverage than History when measured by both lists. This revealed that TED Ed may be a more academic speech event compared to other TED talks, which may provide a suggestion for TED Ed's pedagogical usage in EAP classes.

Several limitations exist in the current study that can provide suggestions for future studies. First, the sizes of the five sub-corpora were not well-balanced; the size of the literature corpus was evidently smaller than other corpora, which may result in the inaccuracy of lexical coverage. Second, only five specific subjects were investigated. The lexical coverage of TED talks concerning other humanities disciplines is unknown, and whether they are also suitable materials for humanities students to learn specialized spoken vocabulary is unsure. Moreover, this study was conducted from a discipline perspective, without examining the influence of TED events on lexical coverage. Future research could further investigate this issue. Finally, it should also be noted that the potential of TED talks for vocabulary learning purposes is not determined by lexical coverage. Other factors, such as collocation, encounters, lexical density, and speech rate, can also make effects. Hence, further research should investigate this issue from more perspectives to promote a comprehensive knowledge of the value of TED talks for vocabulary learning.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author disclosed no potential conflict of interests or personal relationships with respect to the research, authorship, and publication of this article.

**Acknowledgments:** The outcome of this study should thank Prof. Dong who provided me with constant instruction on academic writing, Prof. Chen and his assistant who helped me to overcome the research problem, and other contributors to my study.

## **References**

- [1] Cassidy, R. S., Mike, T., Vincent, L., Andrew, T., Philippe, M., & Benoit, M. (2013). Scientists popularizing science: Characteristics and impact of TED talk presenters. *PLoS One*, 8(4), e62403. <https://doi.org/10.1371/journal.pone.0062403>.
- [2] Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213. <https://doi.org/10.2307/3587951>
- [3] Dang, T. N. (2018a). A hard science spoken word list. *ITL - International Journal of Applied Linguistics*, 169(1), 44–71. <https://doi.org/10.1075/itl.00006.dan>
- [4] Dang, T. N. (2018b). The nature of vocabulary in academic speech of hard and soft sciences. *English for Specific Purposes*, 51, 69–83. <https://doi.org/10.1016/j.esp.2018.03.004>
- [5] Dang, T. N. (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related TV programs: Insights from medical corpora. *TESOL Quarterly*, 54(2), 436–459. <https://doi.org/10.1002/tesq.552>
- [6] Dang, T. N., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66–76. <https://doi.org/10.1016/j.esp.2013.08.001>
- [7] Dang, T. N., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 67(4), 959–997. <https://doi.org/10.1111/lang.12253>
- [8] De Chazal, E. (2014). Prepare English language students for academic listening. British Council. [britishcouncil.org](http://britishcouncil.org). Available at: <https://www.britishcouncil.org/voices-magazine/prepare-English-language-students-academic-listening>. (Accessed 9 April 2022).



- [9] Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3), 157-169. <https://doi.org/10.1016/j.esp.2009.02.002>.
- [10] Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- [11] Humanities. (2022). In *Encyclopædia Britannica*. Retrieved from <http://academic.eb.cnpeak.com/levels/collegiate/article/humanities/41479> (Accessed 20 April 2022).
- [12] Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, 41(2), 235-253. <https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>.
- [13] Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42–53. <https://doi.org/10.1016/j.jeap.2016.01.008>
- [14] Liu, C.-Y. (2021). Examining the implementation of academic vocabulary, lexical density, and speech rate features on OpenCourseWare and MOOC lectures. *Interactive Learning Environments*, 1–16. <https://doi.org/10.1080/10494820.2021.1987274>
- [15] Liu, C.-Y. (2022). Are TED talks potential materials for learning specialized vocabulary? A case of medical vocabulary. *English Teaching & Learning*. <https://doi.org/10.1007/s42321-021-00105-2>
- [16] Liu, C.-Y., & Chen, H. H.-J. (2019). Academic spoken vocabulary in TED talks: Implications for academic listening. *English Teaching & Learning*, 43(4), 353–368. <https://doi.org/10.1007/s42321-019-00033-2>
- [17] Nation, I. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- [18] O'Flynn J. An economics academic word list (EAWL): Using online resources to develop a subject-specific word list and associated teaching-learning materials. *Journal of academic language and learning*. 2019;13(1):A28-A87.
- [19] Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(04), 577. <https://doi.org/10.1017/s0272263109990039>
- [20] Sugimoto, C. R., Thelwall, M., Larivière, V., Tsou, A., Mongeon, P., & Macaluso, B. (2013). Scientists popularizing science: Characteristics and impact of TED Talk presenters. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0062403>
- [21] Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, 12(4), 248–263. <https://doi.org/10.1016/j.jeap.2013.07.001>
- [22] Van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>
- [23] Webb, S., & Nation, P. (2012). Computer-assisted vocabulary load analysis. *The Encyclopedia of Applied Linguistics*. <https://doi.org/10.1002/9781405198431.wbeal0179>
- [24] Wingrove, P. (2022). Academic lexical coverage in TED talks and academic lectures. *English for Specific Purposes*, 65, 79–94. <https://doi.org/10.1016/j.esp.2021.09.004>