



# 英語教材としてのTED Talksの語彙頻度プロファイル分析

杉森, 直樹

---

(Citation)

Journal of Corpus-based Lexicology Studies, 2:1-12

(Issue Date)

2020-03-15

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81011988>

(URL)

<https://hdl.handle.net/20.500.14094/81011988>



## 英語教材としての TED Talks の語彙頻度プロファイル分析

杉森 直樹(立命館大学)  
naosugi@is.ritsumeikan.ac.jp

### Analyzing the Word Frequency Profile of TED Talks as English Learning Materials SUGIMORI Naoki (Ritsumeikan University)

#### 概要

近年, 英語プレゼンテーションの教材として TED Talks が用いられてきているが, 日本人英語者にとっての学習教材としての適切性を語彙指導の観点から考慮した場合, TED Talks にはどの程度のレベルの語彙がどの程度の割合で使用されているのかについての研究が必要であると考えられる。本研究は, 代表的と考えられる TED Talks を 134 本選定し, そのトランスクリプトのデータをコーパス化することにより使用語彙の頻度プロファイル分析を行ったものである。語彙レベル分析の基準として, 日本人英語学習者用の 8000 語の基本語リストである「新 JACET8000」(2016)を用いて TED Talks の使用語彙を分析したところ, リストの 8000 語で 94%のカバー率があり, TED Talks では, トピックによっては専門用語が使用されることがあるものの, 全体的な使用語彙のレベルは学術英語ほどには高くないことが示唆された。また, 対数尤度比検定によって TED Talks の特徴語の分析を行ったところ, *world, human, people*などの語彙や *my, our, me*といった一人称代名詞が特徴語として抽出された。

#### キーワード

TED Talks, 語彙頻度プロファイル, JACET8000

#### 1. はじめに

科学者, 芸術家, 活動家, 思想家, 政治家などが特定のテーマを設定し, 自らの研究や活動, 思想などについてスピーチを行う TED Talks は, Technology, Entertainment, and Design の頭文字を取ったもので, "Ideas worth spreading"をそのミッションとしている。TED は非営利団体として運営され, 話者は原則として 18 分以内のトークによって自らの考えや活動, 研究成果等を述べる。1984 年に会議として発足し, 現在では様々な形態でカンファレンス(Conference)が開催されている。メインとなる TED Conference は年1回アメリカの西海岸で開催される。TEDGlobal はアメリカ以外の国で開催され, これまでイギリスやタンザニア, ブラジル等で開催されている。また, 女性のスピーカーが中心となる TEDWomen や, 若者のための TEDYouth もある。その他にも TED のライセンスを得て個々の団体や学校, 組織等が TED のスタイルでカンファレンスを実施す

る TEDx も毎年世界各地で数多く開催されており、日本でも近年は年間に 50 回以上開催されている。これらの TED conference でなされるトークは TED のウェブサイトでも動画で公開されており、2019 年秋の時点で 3200 以上のトークのビデオが収録されている。扱われるテーマも従来の Technology, Entertainment, Design だけでなく, Business, Science, Global issue など多様なものとなっている。動画には発話のトランスクリプトが用意されており、複数の言語に翻訳したものが用意されているトークも多い。

TED Talks は、スピーチ、プレゼンテーションのみならず、リスニングやリーディング演習のための英語学習教材として利用することが可能であり、英語学習用のテキストも幾つか出版されている (Longshaw & Blass, 2015 など)。その利点としては、実際の英語が素材になっていることや、様々なスピーカーが話す多様な英語の変種に触れられることがある。また、幅広い分野のトピックがあり、具体的なコンテンツがある教材であるため、Content-based Language Teaching (CBLT)や Content and Language Integrated Learning (CLIL)の教材としても利用可能である。更に、TED Talks では映像や音声、実演も含めた多様なメディアでのプレゼンテーションが行われるため、教材としてのリアリティが高いと考えられる。しかしながら、多くの日本人英語学習にとっては、TED Talks を教材としてそのまま利用した場合、使用される語彙の難易度が高すぎて、内容理解が困難になる可能性もあると考えられる。そのため、より効果的に TED Talks を教材として使用するためには、その語彙の難易度や使用される語彙についての計量的な分析が必要と考えられるが、日本人英語学習者にとっての TED Talks の語彙の難易度に関する研究はまだ十分になされていない。そのため、本研究では TED Talks を日本人英語学習者のための教材として捉えた場合に、どの程度の難易度の語彙が使用されているかを調査し、その語彙的特徴をコーパス言語学の立場から分析を行うものである。

## 2. 先行研究

TED Talks は英語学習用の教材として有用であると考えられるため、応用言語学の分野においてもその言語的特徴に関する研究がなされてきている。語彙に関連するものとしては、Coxhead and Walls (2012)は英語の基本語彙リストである General Service List (GSL)と学術語彙リストの Academic Word List (AWL)を用いた TED Talks の使用語彙の頻度分析を行い、他の 3 種類の語彙リストを含めると、92.54%のカバー率があったことを報告している。また、Wang (2012)は TED Talks と BASE (British Academic Spoken English)のコーパスを構築し、使用語彙と 3-gram の頻度分析を行っている。Wingrove (2017)は、アカデミックリスニングにおける TED Talks の適切性についての研究を行い、使用語彙の lexical density と speech rate の分析を行って、TED Talks は多様なアカデミックリスニングに利用可能であると結論付けている。これらの研究は主に海外の英語教育を前提とした研究であるが、日本における英語教育を踏まえた TED Talks の語彙研究はまだ少ない。

### 3. TED Talks における使用語彙の分析

#### 3.1 研究目的

本研究は、TED の公式サイトで公開されているトークのトランスクリプトを収集して独自の TED コーパスを構築し、それを用いて英語学習用の教材としての TED Talks の使用語彙の特徴を分析するものである。前述したように、TED Talk には様々な形態がありスピーカーの特性も多様で、英語の母語話者もいれば非母語話者もいる。英語圏のスピーカーであったとしても国や地域も様々であり、職業分野も異なる。そのため、TED Talks の言語的特徴を全体として分析する場合は、大規模な TED Talks コーパスを構築する必要があるが、英語学習教材としての TED Talks を調査するという本研究の目的を考慮した場合、英語学習用に使用される可能性が高いと考えられる代表的なものを収集してその特徴を分析するという方法を採用の方が適切であると判断した。そのため、本研究では、最も多く視聴されている代表的なトークをまとめたり、テーマ別にトークをまとめたものとして TED が公式サイト等で公表しているプレイリスト(Playlist)に収録されているものを中心にトークの選定を行うこととした。研究のリサーチクエスト(RQ)としては以下の3点を設定した。

RQ1: 日本人英語学習者用の学習語彙リストに基づいて測定した場合、代表的な TED Talks で使用される語彙の難易度やレベル別の分布はどのようなものか？

RQ2: TED Talks を理解するために必要とされる語彙のカバー率に達するには、どの程度のレベルの語彙までを知っていればよいのか？

RQ3: 他の口語英語と比較した場合、TED Talks の特徴語にはどのようなものがあるか？

#### 3.2 TED Talks コーパスの構築

TED Talks のトランスクリプトのデータ収集については、1)その年の最もポピュラーなトークを集めたプレイリスト (2014 年～2018 年)、2)TED Talk 公式ガイドブックのプレイリスト、3)テーマ別に選定されたプレイリスト、に含まれるトークのトランスクリプトを TED のサイトからテキストファイルとして収集、保存した。これらのトークの中には複数のリストに重複して含まれている場合もあったため、それらはどちらかのプレイリストにまとめた。また、長さが 5 分前後の短いトークや 18 分を大幅に超えるような長いトーク、モノログ形式ではなく複数のスピーカーによるダイアログ形式になっているトークについては除外した。選定したトークのリストと数は以下のようにになっている。

表 1 TED Talks コーパスの作成に用いたプレイリストの一覧とトーク数

| プレイリスト                               | トーク数 |
|--------------------------------------|------|
| The Official TED Talk Guide Playlist | 33   |
| The most popular talks of all time   | 17   |

|                                |     |
|--------------------------------|-----|
| The most popular talks of 2014 | 11  |
| The most popular talks of 2015 | 8   |
| Top TED Talks of 2016          | 13  |
| The most popular talks of 2017 | 9   |
| The most popular talks of 2018 | 19  |
| Best science and tech of 2016  | 5   |
| TED MacArthur Grant winners    | 8   |
| 11 must-see TED Talks          | 5   |
| TED Starter Pack               | 6   |
| <hr/>                          |     |
| Total                          | 134 |

最終的に、上記の 11 のリストから合計で 134 本のトークを収集して TED Talks コーパスを構築した。コーパスのサイズは約 33 万語となった。

### 3.3 分析手法

#### 3.3.1 データの事前処理

TED Talks のトランスクリプトには個々の発話のタイムコードや発話者情報が入れられており、発話以外にも(Applause)や(Laughter), (Music)等で表記された会場音の情報も丸括弧でマークされて入れられている。これらのメタデータについては、発話ではないためテキストエディタの正規表現を用いてトランスクリプトから削除する処理を行った。

#### 3.3.2 分析手順

TED Talks コーパスにおける使用語彙の頻度分析については、コーパス分析ツールの *AntConc* を使用して行い、使用語彙の頻度リストを作成した。また、TED Talks コーパスの使用語彙の頻度プロファイル进行分析する基準となる語彙リストとしては新 JACET8000 (2016)を用いた。新 JACET8000 (以下、新 J8) は、JACET8000 (2003)の改訂版として作成されたもので、Corpus of Contemporary American English (COCA)と British National Corpus (BNC)の頻度データをベースとして、中高の英語検定教科書のデータや入試問題のデータ等を補正資料として含めて選定された日本人英語学習者用の 8000 語の学習語彙リストとなっている (リストでは 8000 語が 8 つのレベルに分けられている)。BNC/COCA Word Family Lists (Nation, 2017)や Academic Word List (AWL)などの海外の英語学習用語彙リストでは一般的に word family を基準として語彙が表示されているのに対して、新 J8 は lemma ベースの見出し語で語彙をカウントする形式となっており、日本人英語学習者にとってはより馴染みのある表示方式が採用されている。

本研究ではこの新 J8 を語彙レベルの基準として用いて TED Talks における出現語彙の頻度プロファイル分析を行ったが、分析ツールとしては新 J8 による語彙レベル分析サイト

である J8anWeb を用いた。このサイトでは、任意のテキストに対して新 J8 に基づいた語彙の使用頻度とレベルの分析を行うことが可能で、これにより、TED Talks コーパスの使用語彙の頻度プロファイル进行分析した。

TED Talks の特徴語の分析については、他の口語英語コーパスを参照コーパスとして選定し、それとの比較によって行った。参照コーパスは BNC の Spoken Component のテキストをランダムに収集した BNC Sampler の Context-governed に含まれる Business, Educational, Institutional, Leisure の分野から会話を 51 本収集した。これに加えて academic Spoken English のデータとして MICASE (Michigan Corpus of Academic Spoken English) の講義のトランスクリプトを Physical Sciences & Engineering, Biology and Health Sciences, Social Sciences, Humanities の分野から 22 本収集した。これらを合わせた 70 万語のデータを参照コーパスとした。特徴語の抽出には *AntConc* の Keyness List の機能を使用して TED Talk コーパスと参照コーパスの使用語彙の出現頻度の比較を行った。統計的指標は対数尤度比 (log-likelihood ratio:  $G^2$ ) を使用し Bonferroni 補正を行った。また効果量の指標として特徴語の分析の際に一般的に用いられるオッズ比(odds rate)を使用した。

#### 4. 結果と考察

##### 4.1 RQ1 及び RQ2: 使用語彙のレベル別頻度分布とカバー率

表 2 は *AntConc* を用いて作成した TED Talks コーパスの語彙の頻度リストの上位 50 語である。リストの語を見ると、TED Talks の上位 50 語は一般的な口語英語の高頻度語彙リストと比較的類似しており、*people* が 29 位に入っている以外に大きな特徴は見られなかったが、フィラー(filler)の *erm*, *(u)mm* や *yeah* や *oh* 等の間投詞が入っていない。これは TED Talks は日常会話などの自発的発話(spontaneous speech)とは異なり、事前に用意した原稿に基づいたトークを行っているためであると推測される。

表 2 TED Talks コーパスにおける上位 50 語のリスト

| Rank | Frequency | Word | Rank | Frequency | Word   |
|------|-----------|------|------|-----------|--------|
| 1    | 14268     | the  | 26   | 1748      | can    |
| 2    | 10952     | and  | 27   | 1587      | be     |
| 3    | 9571      | to   | 28   | 1491      | about  |
| 4    | 8147      | is   | 29   | 1488      | people |
| 5    | 7954      | of   | 30   | 1480      | as     |
| 6    | 7575      | a    | 31   | 1465      | all    |
| 7    | 7150      | I    | 32   | 1429      | there  |
| 8    | 7008      | that | 33   | 1404      | at     |

|    |      |      |    |      |         |
|----|------|------|----|------|---------|
| 9  | 5844 | you  | 34 | 1384 | like    |
| 10 | 5275 | in   | 35 | 1371 | do      |
| 11 | 5127 | it   | 36 | 1303 | if      |
| 12 | 4363 | we   | 37 | 1254 | our     |
| 13 | 3380 | are  | 38 | 1242 | or      |
| 14 | 3355 | not  | 39 | 1180 | one     |
| 15 | 3300 | this | 40 | 1147 | your    |
| 16 | 2666 | they | 41 | 1135 | when    |
| 17 | 2459 | so   | 42 | 1107 | me      |
| 18 | 2321 | for  | 43 | 1092 | just    |
| 19 | 2292 | was  | 44 | 1085 | from    |
| 20 | 2197 | but  | 45 | 1003 | an      |
| 21 | 2077 | what | 46 | 992  | how     |
| 22 | 2046 | have | 47 | 978  | out     |
| 23 | 1877 | my   | 48 | 977  | because |
| 24 | 1871 | with | 49 | 977  | now     |
| 25 | 1773 | on   | 50 | 963  | more    |

本研究において収集した TED Talks コーパスの総語数は 332,423 語で、異語数は 16,512 語となった。コーパスのサイズとしては必ずしも大きくはないが、プレイリストのトークを収集したという点において一定程度の代表性を備えていると考えられる。表 3 は、TED Talks の使用語彙が新 J8 の 8 レベルにどの様に分布しているかを示す語彙の頻度プロファイルで、レベル毎の語数(token 数)と総語数に占める割合がパーセントで示されている。また、新 J8 を超えるレベルの語の数(Over 8)や、短縮形(Contracted forms), 数字(Non-words), 固有名詞(Proper nouns)に分類された語の数値も示されている。Level 1 の語が 78.9%と最も割合が高く、その後レベルが上がるにつれて割合が徐々に下がる結果となった。新 J8 は中学校・高等学校で学習すべきと考えられる基本 2000 語とより高いレベルの 6000 語で構成されており、リストの語彙の順位は BNC/COCA のベースリストの順位を日本の英語学習の実態を考慮した補正資料で補正して決定されている。

表 3 TED Talks の新 JACET8000 に基づく レベル別語数と割合

|    | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 |
|----|---------|---------|---------|---------|---------|---------|---------|---------|
| 語数 | 262,245 | 23,608  | 10,360  | 5,631   | 3,690   | 2,283   | 2,014   | 1,474   |
| %  | 78.9%   | 7.1%    | 3.1%    | 1.7%    | 1.1%    | 0.7%    | 0.6%    | 0.4%    |

| Over 8 | Contracted forms | Non-words | Proper nouns | Total   |
|--------|------------------|-----------|--------------|---------|
| 8,579  | 4,235            | 3,212     | 5,092        | 332,423 |
| 2.6%   | 1.3%             | 1.0%      | 1.5%         | 100%    |

図 1 は TED Talks コーパスの新 J8 に基づいたレベル毎の語数(token)とコーパスに対する累積カバー率を示している。カバー率は 8 レベルの語彙のみでカウントした場合と短縮形等を含めた場合との 2 つの数値を折れ線グラフで示した。

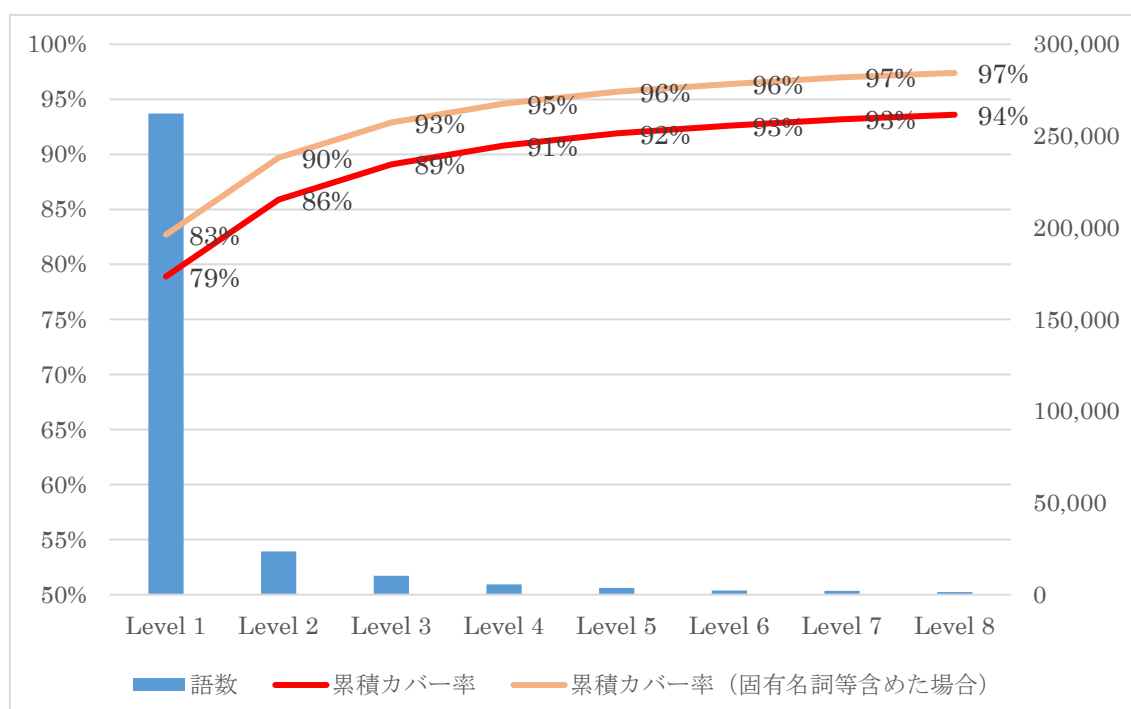


図 1. 新 J8 に基づく TED Talks の使用語彙のレベル別語数と累積カバー率

累積カバー率は Level 4 で 90%を超え、最終的に Level 8 では 94%に達しており、新 J8 の 4000 語までの語彙を知っていれば、TED Talks の使用語彙の 9 割以上はカバーできることが示されているが、これは新 J8 のリストに含まれる語のみでカウントした数値であり、数字や短縮形、固有名詞も理解可能であると仮定してこれらを含めた累積カバー率で見れば Level 8 までで 97%となった。語彙のカバー率と英文理解との関係については、リーディングにおいては 95~98%が必要(Laufer, 1989; Hu & Nation, 2000)とされている一方で、リスニングにおいては Zeeland and Schmitt (2013)は、非母語話者であっても 95%の語彙のカバー率があればリスニングにおいて良好な理解を示したとしている。このことから、固有名詞などの聞き取りに問題が無ければ、新 J8 の 8000 語で TED Talks のリスニング理解に必要な語彙サイズを満たすことができる可能性を示している。また、TED

Talks では豊富な視覚情報が与えられることも考慮すれば、TED Talks のリスニング理解は、考えられているほどには難易度は高くないと言えるかも知れない。

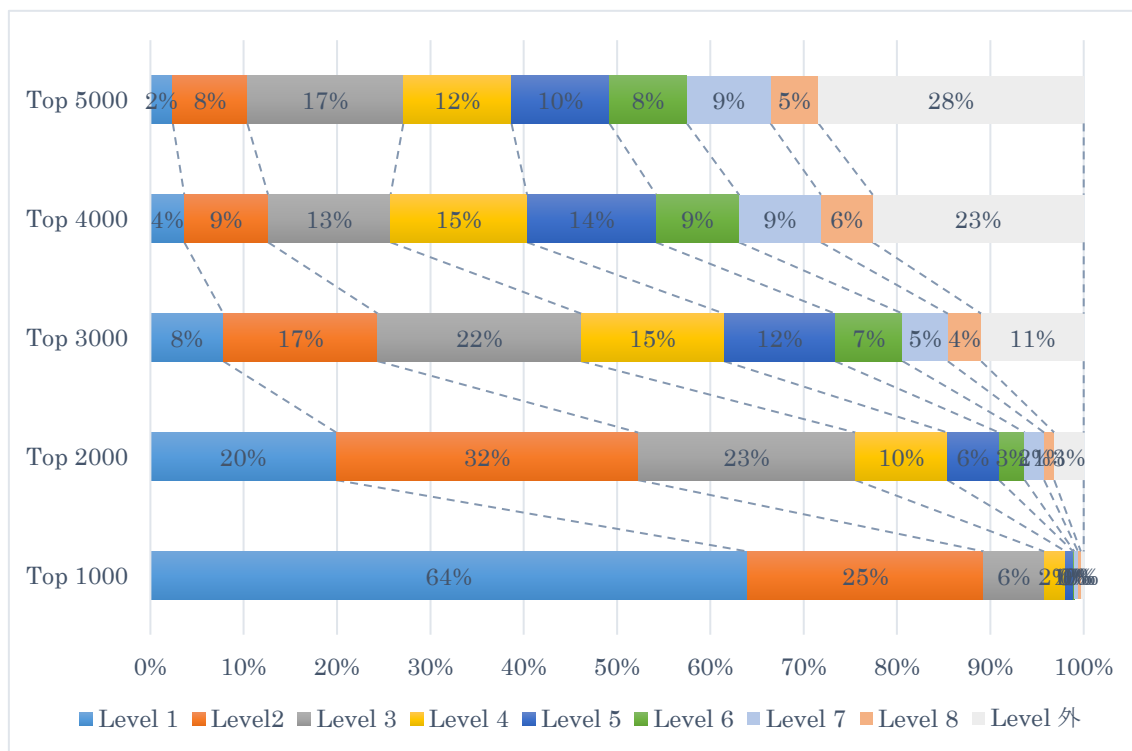


図 2. 上位 5000 語の新 JACET8000 におけるレベル分布の割合

次に、TED Talk コーパスの使用語彙上位 5000 語（実際には同一頻度の語の最後までで切ったため、実際は 5113 語）が、新 J8 ではどのレベルの語彙になっているかを分析した。図 2 は TED Talks コーパスの語彙の高頻度語上位 5000 語を 1000 語毎に分け、それらの語彙が新 J8 のどのレベルに属しているかの比率を示したものである。これによれば、上位 1000 語はほぼ新 J8 の Level 1 と 2 の語彙で占められているが、順位が下がるにつれて新 J8 の上位レベルの語やリストに含まれない語の割合が増える傾向となっており、当然のことではあるが、TED Talks においても出現頻度が下がるほどより難しいレベルの語彙が増える傾向が示されている。しかしながら、Level 3 から Level 5 までの語彙は、TED Talks の上位 3000 語から 5000 語までのそれぞれにおいて常に一定以上の割合で含まれているため、これらの語彙は重要であると考えられる。また、上位 1000 語の中にも新 J8 では高いレベルになる *molecule*(5), *creativity*(5), *complexity*(5), *hemisphere*(5), *compassion*(7) など難しい語彙が含まれている（括弧内の数字は新 J8 におけるレベルを示す）。これらの語はトークのトピックの影響もあると考えられるが、コーパスの用例を確認したところ複数のトークで使用されていたため、TED Talks を教材として使用する際にはこれらの語彙に注意することが必要となるであろう。

## 4.2 RQ3: 特徴語の抽出

表 4 は対数尤度比検定により TED Talks の特徴語として抽出された語彙のリストである。これらの特徴語は主に 2 種類に分類できる。1 つは *brain, life, human, world, people* 等の名詞であり、その使用状況を調査したところ、複数のトークにおいて使用されていたことから、これらは代表的な TED Talks における特徴語であると考えられる。しかしながら、収集したコーパスのサイズの問題から、特定のトークのトピックの影響が一定程度存在する可能性も否定できないため、今後更にコーパスのサイズを増やした上で再分析を行うことが必要であると考えられる。もう一つの特徴語は *my, our, me, us, myself* 等の一人称代名詞である。これについては、TED Talks の手法の一つとして、観客との共感を高める手法としてスピーカーが自分の体験について語る際の *story telling* に使用していると考えられ、TED Talks コーパスの用例を調査すると、”my life” や “my family (parents/father/mother 等を含む)” などのフレーズで多く使用されており、自らの状況について述べたり過去の体験を述べたりする際に使用されているものと考えられる。

表 4 TED Talks の特徴語上位 50 語 (対数尤度比  $G^2$   $p < .05$ )

| Rank | Keyword | Frequency | $G^2$   | Odds ratio |
|------|---------|-----------|---------|------------|
| 1    | my      | 1877      | 1149.66 | 3.50       |
| 2    | brain   | 433       | 702.73  | 18.62      |
| 3    | world   | 647       | 457.02  | 3.99       |
| 4    | our     | 1254      | 382.56  | 2.23       |
| 5    | life    | 506       | 366.16  | 4.09       |
| 6    | human   | 281       | 326.28  | 7.79       |
| 7    | me      | 1107      | 307.88  | 2.13       |
| 8    | people  | 1488      | 227.14  | 1.71       |
| 9    | love    | 200       | 219.00  | 7.02       |
| 10   | and     | 10952     | 204.27  | 1.19       |
| 11   | like    | 1384      | 179.53  | 1.63       |
| 12   | ok      | 103       | 169.62  | 19.72      |
| 13   | stress  | 98        | 163.22  | 20.63      |
| 14   | robot   | 68        | 154.10  | 286.33     |
| 15   | around  | 338       | 150.22  | 2.75       |

|    |           |      |        |        |
|----|-----------|------|--------|--------|
| 16 | universe  | 79   | 150.07 | 41.58  |
| 17 | to        | 9571 | 149.88 | 1.17   |
| 18 | dollars   | 105  | 146.51 | 11.64  |
| 19 | ants      | 72   | 146.33 | 75.79  |
| 20 | us        | 696  | 145.36 | 1.89   |
| 21 | laughter  | 62   | 140.50 | 261.06 |
| 22 | even      | 472  | 136.69 | 2.17   |
| 23 | story     | 181  | 132.80 | 4.14   |
| 24 | years     | 509  | 132.59 | 2.07   |
| 25 | Internet  | 58   | 131.43 | 244.22 |
| 26 | happiness | 75   | 127.57 | 22.56  |
| 27 | lives     | 134  | 127.27 | 5.64   |
| 28 | this      | 3300 | 126.33 | 1.29   |
| 29 | kids      | 121  | 124.78 | 6.37   |
| 30 | planet    | 59   | 117.65 | 62.11  |
| 31 | body      | 198  | 117.02 | 3.39   |
| 32 | learned   | 109  | 116.58 | 6.75   |
| 33 | virus     | 55   | 115.38 | 115.79 |
| 34 | new       | 424  | 114.93 | 2.11   |
| 35 | earth     | 103  | 114.73 | 7.23   |
| 36 | phantom   | 49   | 111.04 | 206.32 |
| 37 | sex       | 92   | 108.87 | 8.07   |
| 38 | myself    | 147  | 107.45 | 4.13   |
| 39 | ant       | 47   | 106.51 | 197.89 |
| 40 | most      | 442  | 103.06 | 1.98   |
| 41 | robots    | 43   | 97.44  | 181.05 |
| 42 | brains    | 52   | 96.88  | 36.49  |
| 43 | every     | 345  | 96.58  | 2.14   |

|    |           |      |       |       |
|----|-----------|------|-------|-------|
| 44 | favorite  | 54   | 96.37 | 28.42 |
| 45 | more      | 963  | 94.69 | 1.52  |
| 46 | power     | 128  | 93.99 | 4.15  |
| 47 | beautiful | 77   | 93.74 | 8.53  |
| 48 | better    | 285  | 93.61 | 2.31  |
| 49 | mom       | 48   | 93.53 | 50.53 |
| 50 | when      | 1135 | 92.02 | 1.46  |

これらの特徴語を見ると、必ずしも学術語彙(academic word)が多い傾向は見られない。Coxhead (2000) によれば、AWL(Academic Word List)は様々な学術分野のテキストにおいて 10%程度のカバー率を達成するとされている。*AntWordProfiler*を用いて TED Talks コーパスの学術語彙を分析したところ、その割合は比較的低く、AWL の語彙の割合は 3.64%であった。TED Talks に含まれる学術語彙の割合に関する他の研究 (Coxhead & Walls, 2012; Wang, 2012) でも同様の低い数値が報告されており、Academic Word List の割合は 2%~3.9%程度であった。このことから、書き言葉と話し言葉の違いはあるにしても、TED Talks は国際会議における学会発表のような学術的なプレゼンテーションよりは使用語彙が易しいものと考えられる。TED Talks のトピックには多様な分野があり、中にはサイエンス系の専門性の高いトークもあるが、全体としてはそれほど多くの学術用語が使用されてはいないことが示されている。その理由については、TED Talks は、専門分野での学会発表とは異なり、一般大衆や専門分野外の素人を対象に、その分野の専門家が分かりやすく説明する popular science (大衆科学) や public speaking の性格を有しているからであると考えられる。TED Talks を教材として使用する際にはこのような特徴を理解しておく必要があると言えるだろう。

## 5. まとめ

本研究においては、小規模なコーパスサイズではあるものの、代表的な TED Talks のデータを収集し、その語彙的特徴についての計量分析を行った。英語学習教材として用いられることも多い TED Talks であるが、使用される語彙に関しては、リスニング理解に必要なカバー率の達成は新 J8 の 8000 語でも可能であることが示唆された。これらのことを考慮すれば、TED Talks の理解には、必ずしも高度な専門語彙(technical vocabulary)の知識は必要ではなく、準専門用語(sub-technical vocabulary)や共通学術語彙を習得させ、TED の特徴語についての知識を与えることで一定の効果を上げられる可能性を示している。また、TED Talks のようなプレゼンテーションを行わせる際には、発表語彙としてこれらの語彙を使いこなせることを目指す指導も重要であると考えられる。

今後、より多くのトークを収集し、今回の分析結果が TED Talk の語彙的特徴としてのどの程度一般化が可能であるかについての研究を行う予定である。

#### 引用文献

- Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp>.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>.
- Coxhead, A., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *TESOLANZ Journal* 20, 55-67.
- 大学英語教育学会基本語改訂特別委員会 (2003). 『大学英語教育学会基本語リスト JACET List of 8000 Basic Words 』 (JACET8000)』 大学英語教育学会.
- 大学英語教育学会基本語改訂特別委員会 (2016). *J8anWeb*. <http://mochvocab.sakura.ne.jp/cgi-bin/j8web/j8web.cgi>
- 大学英語教育学会基本語改訂特別委員会 (編) (2016). 『大学英語教育学会基本語リスト 新 JACET8000』 桐原書店.
- Hu, M. & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13(1), 403-430.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? in Lauren, C. and Nordman, M. (eds), *Special Language: From Humans Thinking to Thinking Machines*. Clevedon: Multilingual Matters.
- Longshaw, R., & Blass, L. (2015). *21st century reading: Creative thinking with TED Talks*. Boston: National Geographic Learning/Cengage Learning.
- Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Available from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457-479.
- Wang, Y. (2012). An exploration of vocabulary knowledge in English short talks: A corpus-driven approach. *International Journal of English Linguistics*, 2(4), 33-43.
- Wingrove, P. (2017). How suitable are TED talks for academic listening? *Journal of English for Academic Purposes*, 30, 79-95. <https://doi.org/10.1016/j.jeap.2017.10.010>.