

令和 3 年 6 月 1 日現在

機関番号：32634  
 研究種目：基盤研究(C) (一般)  
 研究期間：2016～2020  
 課題番号：16K02778  
 研究課題名(和文)TED Talksのスピーチ動画を利用した英日マルチメディア・コーパスの構築  
  
 研究課題名(英文)A Multimedia Corpus for TED Talks  
  
 研究代表者  
 佐藤 弘明(Sato, Hiroaki)  
  
 専修大学・国際コミュニケーション・教授  
  
 研究者番号：60187226  
 交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：良質な英語スピーチが数多く登録されているTED Talks([www.ted.com](http://www.ted.com))の動画3,643作品のセリフをデータベースに登録した。このデータベースを利用して、『新JACET8000語彙リスト』に掲載された8,000語の中の7,949語に対して、単語の意味、英語の用例、用例の日本語訳、用例の書誌情報などを加えたりリストを作成し、それを研究成果として公開した。この語彙リストは、語彙が使用されている場面の動画をウェブブラウザで確認できる。

研究成果の学術的意義や社会的意義  
 本研究では、英単語が使用される場面を動画で確認しながら英語学習が行える電子資料を作成して、それを無料で公開した。これまで公開された英語語彙リストには、発音記号、訳語、英文用例、用例の日本語訳、派生語などは登録されているが、語彙が使用されている場面を動画で確認することはできなかった。本研究で作成した語彙リストには、大学英語教育学会が作成した『新JACET8000語彙リスト』に掲載された英単語が、TED Talksの動画で確認できる。動画付きの英単語リストの語彙は、中学レベルから大学レベルの単語を網羅するため、日本の英語教育に幅広く貢献できる。

研究成果の概要(英文)：Many English vocabulary lists have been published for learners of English in Japan, but most of them do not provide audio-visual data for their vocabulary. This research has developed an audio-visual database for the New JACET8000 vocabulary list, the most popular vocabulary English list in Japan. The database consists of 3,645 titles of TED Talks, and it contains their video clips, transcripts, and bibliographic information. I searched the database for 8,000 lemmas which are listed in the New JACET8000 vocabulary list, and compiled a comprehensive vocabulary list with example sentences and their video clips. This vocabulary list contains 7,949 lemmas, which are essential for English learners in Japan. The list is useful for teaching materials from the basic level to the advanced one.

研究分野：コーパス言語学

キーワード：TED Talks 口語英語 動画 英語 YouTube JACET 語彙 英単語

## 1 . 研究開始当初の背景

膨大な英文電子テキストを収集した British National Corpus や American National Corpus などの均衡コーパスは、語法研究・社会言語学・辞書編纂・文体論などの広い分野に大きな貢献をしてきた。これらのコーパスは、話し言葉のデータも収録してはいるが、その多くはテキストデータしか含まないために、英語の音声や語句が使われる状況を動画で確認することはできなかった。

### 1.1 文字データだけではなく動画を扱う

私は科学研究費補助金研究課題(1-4)で、文字データだけではなく動画を扱うデータベースの研究を行った。

- (1) 英語字幕信号入りレーザーディスクを利用した口語英語データベースの作成 ( 1991 年度 ~ 1992 年度 )
- (2) 口語英語研究のためのマルチメディア型データベースの作成 ( 1994 年度 )
- (3) クローズド・キャプションを利用したデータベースの作成 ( 1995 年度 )
- (4) CC 付きビデオを利用したマルチメディア型データベースの構築 ( 1996 年度 )

これらの研究成果を公開するために、開発したデータベースの一部の著作権処理をして(5)を発表した。

- (5) CC データベース、平成 9 年度文部省科学研究費補助金「研究成果公開促進費」( データベース ) の研究成果、作成組織：映画英語教育学会研究委員会

(1-5) の研究成果により、英語表現をデータベースで検索して、それが使われる場面を動画で確認することが可能となった。しかしながら、動画の再生にはレーザーディスクという古い媒体を使うため、現在、一般の研究者がこのデータベースを使用しやすい状況にはない。現在では、**YouTube** のように**ウェブブラウザで動画が視聴**できるソフトウェアが求められる。英語学の研究では、Google 検索結果を利用する例は多いが、インターネット上の動画のデータを利用したものは極めて少ない。

### 1.2 先行研究

インターネット上の動画を英語コーパスとして使用した先行研究としては、Hasebe(2015)がある。この研究では、良質な英語スピーチが数多く登録されている TED Talks (<http://www.ted.com>)スピーチ動画から任意の語句を検索し、その語句が使われている場面を Web プラザー上で再生するシステム(TCSE, <http://yohasebe.com/tcse>)を公開している。TED Talks の動画とテキストは、幅広い利用を認めるクリエイティブ・コモンズ・ライセンスという著作権を採用しているため、TCSE のようにデータベース化して一般公開することが可能である。TCSE は、多くの研究者に利用されており高い評価を得ている。本研究の主な目的は、TCSE を英語語彙学習などの英語教育に応用するものである。

## 2 . 研究の目的

### 2.1 動画付き語彙リストの作成

大学英語教育学会(JACET)では、1981 年より日本の英語学習者が学ぶべき基本英語語彙リストを作成してきた。その最新版が『新 JACET8000 語彙リスト』である。この語彙リストの基本項目は、単語のレベル番号、単語の基本形であるレマ(lemma)、品詞という 3 つから構成される。例えば、表 1 はレベル番号 2001~2005 までのリストである。

J レベル番号	レマ	品詞
2001	arrange	動詞
2002	recommend	動詞
2003	flour	名詞
2004	oxygen	名詞
2005	gene	名詞

表 1 : 新 JACET8000 語彙リスト

『新 JACET8000 語彙リスト』には単語の意味や英語の用例がないため、英語学習者がこのリス

トをそのまま使用して、英単語を学習することは難しい。この問題を解決するために、本研究では、TED Talks から収集したデータを利用して、表 1 のようなデータに単語の意味、英語の用例、用例の日本語訳、用例の書誌情報などを加え、図 1 のようなリストを作成する。図 1 は html ファイルであり、ウェブブラウザで閲覧できる。それぞれの見出し語には、「文脈」と「解説」と書かれた 2 つのボタンがある。「文脈」をクリックすると、「見出し語の用例が使われている場面の動画が視聴できる。この図では、arrange の項目の中の「文脈」をクリックして、表示されている用例 “and here I am trying to ^arrange things.” が使用されている場面の映像を再生している。

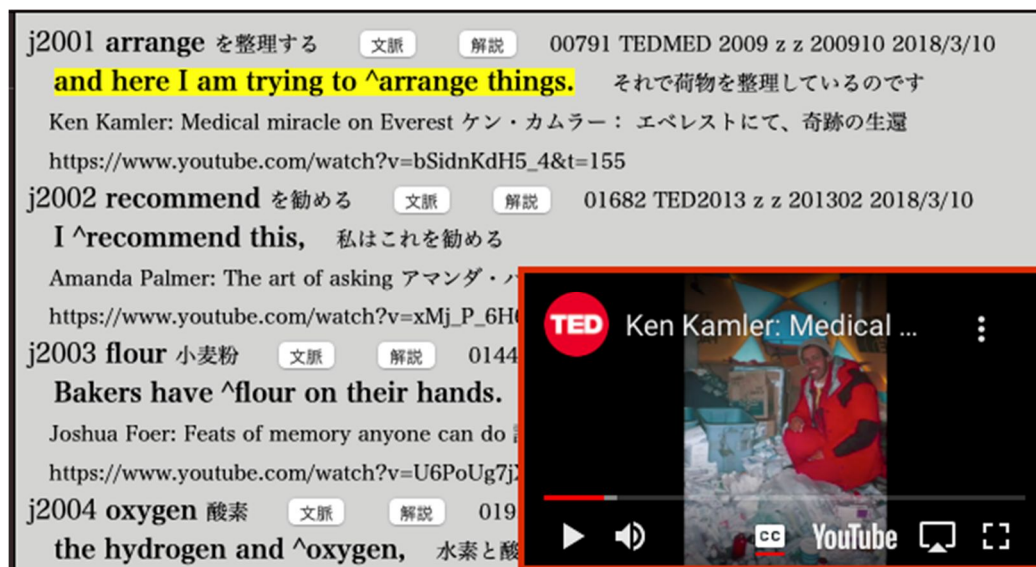


図 1：TED Talks から用例などを追加した動画付き語彙リスト

### 3. 研究の方法

#### 3.1 データの収集と登録

TED.com では、TED Talks のスピーチ動画の字幕データがダウンロードできる。通常は、ダウンロード用のリンクをマウスでクリックして字幕データをダウンロードするが、本研究ではコンピュータ言語 python で作成したプログラムを利用して、大量の字幕データを一括してダウンロードしている。また、スピーチのタイトル、講演者、要旨などの情報も一括してダウンロードするプログラムも作成した。ダウンロードしたデータをデータベースに登録し、検索した英語語句が使われている場面のビデオ映像をウェブブラウザ上で確認できるデータベースを作成した。

このデータベースを利用して『新 JACET8000 語彙リスト』の語彙を検索して、用例を登録する作業をした。用例は動画ファイルとリンクしているため、用例が使用されている場面の音声と映像を確認することができる。

#### 3.2 語彙データ入力方法

表 1 のような『新 JACET8000 語彙リスト』から、図 1 のような html 形式の語彙リストを作成するために、語彙リスト登録用ソフトウェア（以下、新 JACET8000 登録ソフト）を開発した。

新 JACET8000 登録ソフトでは、検索作業の多くをマウスのクリックで実行できるようにした。先行研究の TCSE では、検索する英単語をキーボードから文字入力したが、新 JACET8000 登録ソフトでは、文字入力の代わりにマウスのクリックで英単語の検索を実行する。例えば、図 2 の右下は、表 1 の英単語を検索するためのメニューが表示されている。このメニューで各単語の右のボタンをクリックすると、その単語が TED Talks で使用されているセリフを検索して、検索結果を表示する。図 2 では、左側に検索した単語を含む TED Talks 内の英文、右側にそのに対応する日本語字幕が表示される。

データ登録では、検索結果の中から適切な例文を選び出し、それをクリックする。図 2 では 9 番目の例文がクリックされている。例文をクリックすると、YouTube 上の TED Talks 動画がウェブブラウザ上で再生される。このビデオを見ながら、音声が聞き取りやすいものを用例として選ぶようにした。

図 3 は、データ入力用の画面である。この画面でも、入力の効率化のために、できる限りキーボードからの手入力を避けて、マウスのクリックでデータ入力をするようにした。例えば、図 2 のように 9 番目の例文をクリックすると、図 3 のデータ入力画面の多くの欄に、見出し語 arrange、英語の用例 “and here I am trying to ^arrange things.”、それに対応する日本語字幕のデータが自動入力される。

1	The other things that religions do is to ^arrange* time.	宗教はまた 時間を設定します	
2	and we ^arrange*d ourselves in front of one of the barricades,	バリケードの正面に陣取りました	
3	arranging it the way we wanted?	自由にアレンジできないかと	
4	already in an ^arrange*d marriage,	決められた相手と結婚して	
5	and we ^arrange* them in an image.	これらを画像として並べ	
6	so thanks for arranging for that.	お手配ありがとうございます	2001 arrange/動詞 <input type="radio"/>
7	Orchestral arranging is constructive.	管弦楽の編曲は構成的で	2002 recommend/動詞 <input type="checkbox"/>
8	I have ^arrange*d them in a full house.	フルハウスを仕込みまし	2003 flour/名詞 <input type="checkbox"/>
9	and here I am trying to ^arrange* things.	それで荷物を整理してい	2004 oxygen/名詞 <input type="checkbox"/>
10	So -- I ^arrange* the spades -- the clubs.	スペードを揃えます そし	2005 gene/名詞 <input type="checkbox"/>

図 2：動詞 arrange の検索結果一覧

図 3：データが自動入力された画面

図 3 では、arrange の訳語である「を整理する」は自動入力されていないので、それをキーボードから入力して登録するれば、図 1 の arrange の項で表示されているデータ入力完了する。図 3 では例文の出典を示す書誌情報や動画の URL が表示されていないが、実際には非表示の欄に自動入力されているため、キーボードから手入力する必要はない。また、図 3 のボタンには、複数の英和辞書を参照できる「辞郎」や「内辞」ボタンなどデータ入力を円滑に行うための機能がある。

#### 4. 研究成果

「3.1 データの収集と登録」の作業の結果、2021 年 5 月時点で TED Talks の動画 3,643 本、総英単語数 7,612,874 語をデータベースに登録した。そのデータベースを検索して、『新 JACET8000 語彙リスト』の 8,000 語を検索し、適切な場面が検索できた 7,949 語に関して、「2.1 動画付き語彙リストの作成」で述べた項目を入力して、動画付きの語彙リストを(6,7)の URL で公開した。

(6) <http://sato.fm.senshu-u.ac.jp/TED/zPublic/index.html>

(7) <http://sato.fm.senshu-u.ac.jp/TED/zPublic/001abcList.html>

(6)は、図 4 のような語彙リストを『新 JACET8000 語彙リスト』の語彙番号順に 100 語ごとに区切った索引である。

1.	<a href="#">0001-0100</a>
2.	<a href="#">0101-0200</a>
3.	<a href="#">0201-0300</a>
4.	<a href="#">0301-0400</a>
5.	<a href="#">0401-0500</a>
6.	<a href="#">0501-0600</a>
7.	<a href="#">0601-0700</a>
8.	<a href="#">0701-0800</a>
^	<a href="#">0801-0900</a>

図 4：100 語ごとに区切った語彙リストの索引

図4の8.0701-0800のリンクをクリックすると、『新 JACET8000 語彙リスト』の語彙レベル701-800の英単語のリストが図5のように表示される。この図では、見出し語 accident の **文脈** ボタンをクリックして、この用例が使用されている TED Talks の動画を再生している。



図5：701-800の英単語のリスト

(7)は、『新 JACET8000 語彙リスト』の語彙をアルファベット順に並べた図6のようなリストである。単語をクリックすると、その単語が使用されている場面を動画で確認できる。リンクが設定されていない winger, ye, zest などは英単語については、TED Talks 内での適切な使用例が存在しなかった英単語である。

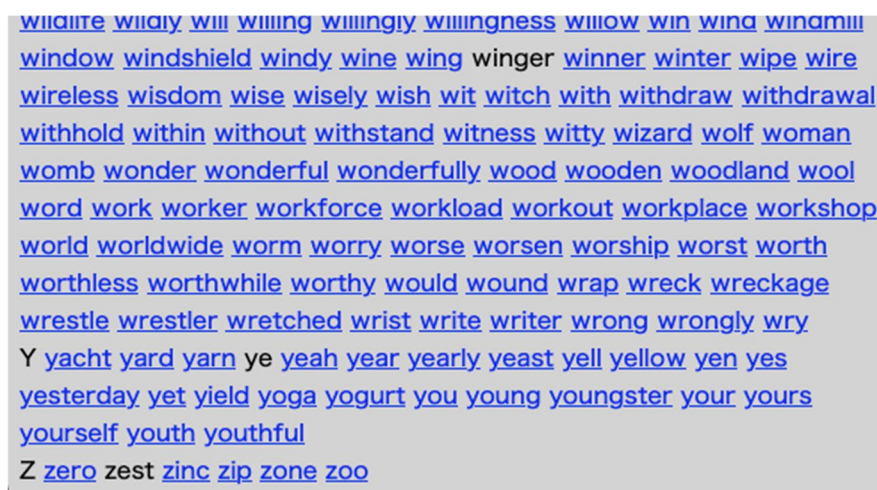


図6：アルファベット順の語彙リストの索引

これまで公開された英語語彙リストには、発音記号、訳語、英文用例、用例の日本語訳、派生語などは登録されているが、語彙が使用されている場面を動画で確認することはできなかった。本研究で作成した語彙リストは、**語彙が使用されている場面を TED Talks の動画で確認できる**点で、研究成果の学術的意義や社会的意義が高い。

#### 参照資料

- 大学英語教育学会基本語改訂特別委員会(編著) (2016) 『大学英語教育学会基本語リスト 新 JACET8000』, 東京: 桐原書店.
- Hasebe, Y. (2015). Design and Implementation of an Online Corpus of Presentation Transcripts of TED Talks. *Procedia: Social and Behavioral Sciences* 198(24), 174-182.

## 5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 0件/うち国際共著 0件/うちオープンアクセス 7件）

1. 著者名 佐藤 弘明	4. 巻 49
2. 論文標題 オンライン授業に対応した英語試験用アプリの開発	5. 発行年 2021年
3. 雑誌名 専修大学外国語教育論集	6. 最初と最後の頁 113-122
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 佐藤 弘明	4. 巻 49
2. 論文標題 オンライン授業に対応した英語試験用アプリの開発	5. 発行年 2021年
3. 雑誌名 専修大学外国語教育論集	6. 最初と最後の頁 113-122
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 佐藤弘明	4. 巻 48
2. 論文標題 字幕ファイルを利用した多言語教育用ソフトウェアの開発	5. 発行年 2020年
3. 雑誌名 専修大学外国語教育論集	6. 最初と最後の頁 93-101
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 佐藤 弘明	4. 巻 47
2. 論文標題 『新JACET8000語彙リスト』に対応したTEDの動画付き用例集	5. 発行年 2019年
3. 雑誌名 専修大学外国語教育論集	6. 最初と最後の頁 101-114
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 佐藤弘明	4. 巻 46
2. 論文標題 TED Talksの英日字幕データを利用した語彙リストの作成	5. 発行年 2018年
3. 雑誌名 専修大学外国語教育論集	6. 最初と最後の頁 105-115
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 佐藤弘明	4. 巻 17
2. 論文標題 各学生の学力に合わせた英語試験の実施	5. 発行年 2017年
3. 雑誌名 教育開発支援NEWSLETTER	6. 最初と最後の頁 3-4
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 佐藤弘明	4. 巻 44
2. 論文標題 TEDビデオを利用した英日マルチメディア・コースの構築	5. 発行年 2017年
3. 雑誌名 専修大学外国語教育論集	6. 最初と最後の頁 101-112
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

#### 6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

#### 7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------