

論文

TED Talks で使用される Lexical Bundles の
特徴に関するコーパス研究

杉 森 直 樹*

要旨

In the field of English-language teaching, TED Talks have been increasingly utilized as authentic materials for instructing various skills such as academic listening, oral presentation and public speaking. Despite considerable research interest in the linguistic features of this unique genre of spoken English, only a few studies have attempted to investigate the lexical and discursive features of the talks. This corpus-based study aims to investigate the characteristics of lexical bundles used in TED Talks and reveal their discourse functions. The transcripts of 115 popular TED Talks were collected from the official TED website to construct a mini TED Talk corpus, and frequency lists of three- and four-word lexical bundles were obtained using the corpus analysis tool *AntConc*. Log-likelihood tests were conducted to extract key lexical bundles which occur more frequently in the TED Talk corpus than in the reference corpus. The results of the test identified the key lexical bundles which were used more frequently in TED Talks. Furthermore, functional classification of 4-word key lexical bundles was attempted to investigate their discourse functions in the talk. The results suggested that TED Talk speakers tend to employ particular lexical bundles to express their desire or intention in the talk so that they can effectively disseminate their ideas to the audience in an organized manner.

キーワード

lexical bundles, TED Talks, discourse functions

* 立命館大学情報理工学部 教授

目 次

- I. はじめに
 - 1. Lexical bundles の定義
 - 2. TED Talks における lexical bundles
 - 3. TED Talks のデータ収集法
- II. TED Talks コーパスの収集と分析方法
 - 1. Lexical bundles の抽出
 - 2. 参照コーパスとの頻度比較
- III. 結果と考察
- IV. まとめ

I. はじめに

本研究は、オーセンティックな英語プレゼンテーションの教材として近年注目されてきている TED Talks における lexical bundles（単語連鎖）の使用に関して、コーパス言語学の立場から計量的分析を行い、TED Talks で使用される lexical bundles の特徴の解明を目的とするものである。リスニングやプレゼンテーションの授業の教材として使用されることも多い TED Talks であるが、そこで使用されるフレーズの特徴についてはまだ十分な研究が行われておらず、どのようなフレーズを重要なものとして指導を行うかや、それらがトークの中でどのような機能を果たしているのかを学習者に説明するためには、これらの使用実態の解明を行う必要がある。そのため、本研究では TED Talks のミニコーパスを作成し、それをを用いて高頻度で使用されるフレーズを lexical bundles として分析するものである。

1. Lexical bundles の定義

英語では隣接または近接する語がまとまってフレーズとして定型的に使用されるものがあり、一般には formulaic language/sequences や multiword units などと呼ばれている。また、これらに類似したものとして formula, chunks, lexical phrase, prefabricated phrases/routine 等があり、idiom や collocation も含め、このようなフレーズ表す用語が数多く用いられている (Schmitt, 2004)。このようなフレーズで構成される定型表現は、多くの場合心理的にはひとまとまりとなって処理されるとされており、それが用いられているジャンル (genre) やレジスター (register : 言語使用域) の特徴を示す指標の一つとして近年関心が高まってきている。

しかしながら、これらの用語はフレーズのイディオム性や意味の透明性 (transparency) の有無等によってその定義が異なり、あるフレーズがそれに該当するかの判断を機械的に行うことは容易ではない。計量的分析によってフレーズ研究を行う場合、このような問題を解決するための用語として、lexical bundles (Biber et al, 1999) がある。Lexical bundles は、テキスト

の中に高頻度で出現する 3 語以上の単語連鎖であり、コーパスからの n-gram (n 個の語の連鎖) 解析によって頻度情報と共に抽出が可能であるため、フレーズ研究でしばしば問題となるイディオム性の有無やフレーズ構造の判別にとらわれないことが特徴である。そのため、lexical bundles はコーパス言語学の分野での計量分析による研究に適した概念であると言える。

Biber et al. (1999) では lexical bundles を以下のように定義している。

“Lexical bundles are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse.”

これによれば、lexical bundles には単語 (single word) と同程度の頻度で使用されるものもあるとされており、会話では語彙の 30% が lexical bundles として出現し、学術散文 (academic prose) では 21% の語が lexical bundles として使用されていると述べている。また、idiom とは異なり、*it was in the* のようなフレーズとしては不完全な形の語連鎖であっても lexical bundles に含まれるとされており、また、短縮形 (*don't* など) を一語としてカウントするのも lexical bundles の特徴となっている。

基本的には lexical bundles は機械的な語の連鎖であるが、出現頻度とテキスト間の分布度合いの観点から、その語連鎖を lexical bundles とするかどうかについては以下の基準が設定されている。

- 1) 4 語の lexical bundles の場合、一つのレジスターにおける出現頻度が 100 万語あたり 10 回以上であること (5 語や 6 語の場合は、5 回以上)
- 2) 少なくとも 5 つ以上の異なるレジスターに出現していること

これらの 2 つの基準の両方を満たしたものが lexical bundles であるとされており、この場合のレジスターはテキストのカテゴリーを意味する場合もある。

他の定型表現と同様に、lexical bundles もレジスターやディスコースの特徴を示す指標であるとされており、会話と学術散文では使用頻度の高い lexical bundles が異なるとされている。例えば、会話では *I don't know what* などの personal pronoun + lexical verb phrase のパターンが多く使用される傾向があるのに対して、academic prose では *the nature of the* のような noun phrase with pose-modifier fragment のパターンや、*as a result of* のような preposition + noun phrase fragment のパターンが多いとされている。これらのことから、あるテキストにおいて高頻度で使用される lexical bundles を分析すれば、そのテキストがどのレジスターとしての特徴を持っているかを分析することが可能であると言える。

2. TED Talks における lexical bundles

TED Talks は “Ideas worth spreading” をそのミッションとして、科学者や芸術家、思想家などが自らの研究や作品、思想などについてパブリックスピーチを行うイベントであり、Technology, Entertainment, and Design の頭文字を取って TED となっている。1984 年に始まり、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。TED Conference は年 1 回アメリカで TED 主催で開催されるが、個々の団体や学校、組織等がライセンスを得て独自の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いる。

TED では Technology, Entertainment, and Design だけでなく、Business や Science, Global issue などの分野の専門家も一般の聴衆に対してステージ上での実演を含めて様々なメディアを用いてプレゼンテーションを行っている。そのため、言語的には単にパブリックスピーチという分類だけではなく独自のジャンルを構成していると言え、そこで使用される英語に対する研究も行われてきている。また、英語教育の分野では TED を英語プレゼンテーションの教材として使用することも増えてきており、TED をベースとした英語のテキスト (Longshaw & Blass, 2015) も出版されている。日本でも NHK の E テレで「スーパープレゼンテーション」というタイトルの番組として放送されていたため学習者の関心も高く、リスニング教材やプレゼンのモデルとして使用されている。

本研究は TED Talks で使用される lexical bundles について、独自に収集したコーパスを用いた分析を行い、lexical bundles の頻度データを得ると共にその用法について考察を行うことを目的とするもので、Research question としては以下の 3 つを設定した。

- 1) TED Talks において高頻度で使用される lexical bundles はどのようなものか？
- 2) TED Talks の特徴的フレーズとして使用される lexical bundles はどのようなものか？
- 3) TED Talks で特徴的に使用される lexical bundles はどのような談話機能 (discourse function) を持っているか？

TED Talks で用いられる lexical bundles に関する研究については、Wang (2012) が Science & Technology, Global issues, Business の分野の TED Talks のスピーチを約 8 万語集めたコーパスをもとに分析を行い、3 語の lexical bundles のリストを作成した。Shin and Shin (2020) は global issues に関連する貿易関連のテーマの 30 本の TED Talks を集めた約 6 万語のコーパスと国際貿易関連の BBC ニュースを集めた 2 つのコーパスを作成し、両者における lexical bundles の談話機能や統語構造の違いを比較している。また、Liu and Chen (2020) は、lexical bundles の果たす談話機能を academic lecture と TED Talks で比較し、TED Talks では discourse organizer の役割を果たす lexical bundles の使用が多いことを報告している。これらの先行研究が示すように、TED Talks においても lexical bundles はそのジャンルの特徴を示す指標となっていることが示され、TED Talks における lexical bundles の研究

の重要性が高まっている。

3. TED Talks のデータ収集法

前述したように、TED Talks は多様な形態で開催され、その話者の特性（国籍や母語、職業、ジェンダー、年齢等）も多様である。そのため、TED Talks の特徴をコーパスを用いて計量的研究を行う場合、二つのアプローチがあると考えられる。一つは大量にデータを収集し、TED Talks 全体としての特徴を分析するものである。このアプローチは大量データの長所を生かして俯瞰的な分析が可能になりデータの信頼性も高くなる一方で、Web サイトからの大規模なスクレイピング（scraping）等で定期的にデータ収集を行わなければならないという問題があることや、TED Talks の各種イベントの違いや話者特性の違いを考慮せず、TED Talks 全体として分析を行う形になる。研究成果を TED Talks を用いた英語教育に応用する場合はこの点に注意が必要であり、このようなアプローチによる分析結果を見る場合は多様な TED Talks を集めた全体的特徴として捉える必要があると言える。

もう一方のアプローチは、研究結果を英語教育に応用することを主たる目標として、教科書や教材として英語の授業で利用されるような代表的、典型的な TED Talks を収集し、その特徴を解明する形である。この場合には収集するデータの量は制限されるが、多様なバリエーションがある TED Talks のうち、英語学習者の教材として適している教科書的なトークの持つ特徴を分析することになるため、教材として TED Talks を利用する場合に役立つデータが得られるという利点がある。本研究ではパイロットスタディであることも考慮し、後者のアプローチを取ることとし、代表的な TED Talks を TED の公式 Web サイト（TED.com）から収集することとした。

II. TED Talks コーパスの収集と分析方法

TED Talks の Web サイトには 2020 年 10 月現在で 3,600 を超える TED Talks が収録されており、トークのビデオを視聴することができだけでなく、スピーチの transcript を読むことができる。そのため、本研究では transcript をテキストデータとして収集して TED Talks のミニコーパスを作成することとし、代表的な TED Talks を収集する際の選定基準として、TED.com が作成している Playlist（特定のテーマや分野のトークをまとめた視聴リスト）を利用することとした。TED.com には様々な Playlist が掲載されているが、TED Talks としての代表性の高いものを選定するという目的を考慮し、*The most popular 25 talks of all time*, *TED Starter Pack*, *The Official TED Talk Guide Playlist*, *The most popular talks of 20XX* 等の多くの人が視聴する代表的なトークを集めた Playlist を選び、そこに含まれているトー

クのスクリプトをテキストファイルで個々に収集して TED Talk のコーパスを構築した。TED がそのガイドラインで規定しているトークの時間が基本的には 18 分までであることを考慮し、これより大幅に長いものや短いものは除外した。また複数の Playlist 間で重複しているものについてはどちらか一つを採用した。

最終的に本研究で使用するミニコーパスとして収集したトークの本数は 115 本で、総語数 (token) は 289,508 語、異語数 (type) 15,575 語で、1 トークの平均語数は 2,518 語であった。なお、transcript 中の発話のタイムコードや発話者情報、(Applause), (Laughter), (Music) 等の会場音の情報等のメタデータは削除している。

1. Lexical bundles の抽出

収集した TED Talk コーパスのテキストファイルをコーパス分析ツールである *AntConc* を用いて分析した。*AntConc* は多様なコーパス分析機能を備えており、読み込ませたテキストファイルに対して、語彙の使用頻度をカウントして頻度表を作成する機能や任意の語彙のコンコードダンス (KWIC) 表示、語彙のテキスト中での出現位置の表示、n-gram 解析、コロケーションの分析などを行うことができる。本研究では、最初に使用語彙の頻度カウント機能を使用してコーパスサイズの確認を行い、引き続いて TED Talks で使用されている lexical bundles を抽出するために n-gram 解析の機能を使用して 3-gram と 4-gram の lexical bundles の頻度カウントを行った。なお、前述したように、lexical bundles に認定する基準としては、出現頻度が 100 万語あたり 10 回以上あり、5 つ以上のトークに出現しているものという基準を採用した。

2. 参照コーパスとの頻度比較

TED Talks の特徴フレーズの分析には、コーパスの n-gram 解析によって得られた 3 語及び 4 語の lexical bundles が TED Talks だけでなく口語英語全般においても一般的に高頻度で利用されるものであるか、或いは TED Talks において特徴的に高頻度で利用されるものであるかを判別することが必要である。ある特定のコーパスにおいて他の一般的なコーパスより高い頻度で利用される傾向がある語彙やフレーズは特徴語 (keyword) や特徴フレーズ (keyphrase) と呼ばれる。これらの特徴語を分析する手法としては、より一般的な大規模コーパスを参照コーパス (reference corpus) として用い、調査対象のコーパスとの出現頻度の比較を行い、2 つのコーパスにおける出現頻度の差を対数尤度比検定 (log-likelihood test) や χ^2 検定 (Chi-squared test) によって検証する手法がとられる。

本研究では、収集した TED Talks コーパスとの比較対象として British National Corpus (BNC) を参照コーパスとして利用することとした。BNC は現在では様々な形態で提供され

ているが、本研究ではその一部を収録した BNC Sampler の口語英語のデータである Spoken Component の Context-governed のデータを参照コーパスとした。Sampler の Context-governed の口語英語のデータは全部で 51 本のテキストが収録されており、Educational/Informative, Business, Public/Institutional, Leisure の 4 つの register のデータで構成され、総語数は約 50 万語となっている。TED Talks の n-gram 解析によって得られた 3 語及び 4 語の lexical bundles の個々の出現頻度を BNC におけるそれらの出現頻度と比較し、対数尤度比検定を用いて 2 つのコーパス間での出現頻度に有意差があるかを検証した。

III. 結果と考察

AntConc を用いた lexical bundles の頻度分析を行った結果、3 語の lexical bundles が 2,219 個、4 語の lexical bundles が 288 個抽出された。それぞれの上位 100 を付表 1 に示す。

3 語の lexical bundles の頻度表を見てみると、*a lot of*, *I want to*, *when I was*, *you can see* などの口語英語的な表現が上位にランクされていることが示されている。また、Biber et al. (1999) でも示されているように、3 語の lexical bundlesの中には 4 語の lexical bundles の一部になっているものも存在している。4 語の頻度リストでは *if you look at* や *it turns out that* などが上位に来ている。

これらの表現が TED Talks で多く用いられることについては、TED Talks ではスピーカーが自らの体験を話す *storytelling* のスタイルを取ることが多いことや、その分野の専門家が一般の聴衆に専門的な内容を分かりやすく解説する *science popularization* の要素や *dissemination of knowledge* の特性を持つことが現れていると考えられ、TED Talks の使用フレーズを特徴付けるものの一つとなっていると言えるであろう。

表 1 は先行研究が示している TED Talks の 3 語の lexical bundles の上位 20 を本研究の結果と比較したものであるが、異なる TED Talks を分析しているにも関わらず、*a lot of*, *this is a*, *we have to* などの lexical bundles が 3 つのリストに共通して高頻度で出現していることが分かる。これらの表現は TED Talks に共通して使用頻度が高いと判断される。

TED Talks の言語的特徴の一つとして、スピーカーが自己の主張を聴衆にアピールする手法として近接性 (proximity) を使用することが多いとされている (Scotto di Carlo, 2014)。これは一人称の代名詞である *we* や *I* を含む表現などによって実現され、スピーカーがこれらの語を使用することで聴衆の共感を得て心理的距離を縮め、トークの説得力やアピール力を増そうとする手法である。高頻度の lexical bundles を見ると、このような手法が使用されていることが分かる。

本研究において抽出した lexical bundlesの中には、TED Talks においてだけでなく口語英

表 1 3 語の lexical bundles の出現頻度順位の比較

Rank	present study	Wang (2012)	Shin and Shin (2020)
1	a lot of	a lot of	around the world
2	I'm going to	one of the	it is not
3	one of the	this is a	this is a
4	I want to	there is a	in the world
5	this is a	a little bit	one of the
6	this is the	## percent of	of the world
7	and this is	you can see	tell you that
8	a little bit	we have to	a lot of
9	going to be	we need to	in war zones
10	when I was	the United States	we have to
11	in the world	around the world	we need to
12	you have to	in the world	as a result
13	you can see	a couple of	how to do
14	you want to	the fact that	I want to
15	be able to	you have to	in our cities
16	it turns out	look at the	of the most
17	this is what	part of the	the European Union
18	we have to	some of the	in the south
19	the united states	the kind of	some of you
20	out of the	the rest of	the number of

語一般において使用頻度が高い傾向があるものも含まれると考えられるため、どれが TED Talks における特徴フレーズ (keyphrase) であるかを分析する必要がある。そのため、参照コーパスの BNC の頻度データとの比較を行い、対数尤度比検定を行って TED Talks に特徴的に出現している lexical bundles を分析した結果、126 個の 3 語の lexical bundles と、48 個の 4 語の lexical bundles の使用頻度が TED Talks において有意に高い ($p < .05$) 特徴フレーズであることが判明した。これらの上位 20 のリストを表 2 と表 3 に示す。

これらの表を見ると、*it turns out (that)* や *I want to talk, going to show you, when I was (in)* などの表現が対数尤度比 (G^2) が高いことが示されており、プレゼンテーション形式のパブリックスピーチとしての TED Talks の中でこれらの表現が特徴的に使用される傾向があることが分かる。

更に、これらの特徴的な lexical bundles が TED Talks においてどのような談話機能を持っているかについて 4 語の lexical bundles を用いて分析を行った。なお、3 語の lexical bundles については、どのような談話機能を持つかの判定が困難な場合も多いことから今回は分析を行わなかった。談話機能の分類については、Biber (2006) で示されている spoken academic English における lexical bundles の functional classification の基準を用いた。これによれば、4 語の lexical bundles は、表 4 に示すような談話機能を持つとされている。

表 2 TED Talks における 3 語の特徴フレーズ (keyphrase)

3-word lexical bundles	G^2
<i>it turns out</i>	114.57
<i>of the brain</i>	98.49
<i>I want to</i>	76.70
<i>to figure out</i>	76.38
<i>in the world</i>	74.20
<i>I'm going to</i>	72.51
<i>the United States</i>	71.63
<i>when I was</i>	60.66
<i>and she said</i>	56.28
<i>in the brain</i>	52.26
<i>a way to</i>	50.25
<i>the world and</i>	48.35
<i>turns out that</i>	46.23
<i>a bunch of</i>	44.22
<i>in my life</i>	44.22
<i>to tell you</i>	42.62
<i>this is what</i>	41.17
<i>going to show</i>	40.20
<i>I started to</i>	40.20
<i>oh my god</i>	40.20

表 3 TED Talks における 4 語の特徴フレーズ (keyphrase)

4-word lexical bundles	Discourse Function	G^2
<i>it turns out that</i>	Identification/focus	42.21
<i>I want to talk</i>	Desire-Personal	40.20
<i>going to show you</i>	Intention/prediction	36.18
<i>I'm going to show</i>	Intention/prediction	30.15
<i>part of the brain</i>	Other Referential	30.15
<i>in the United States</i>	Other Referential	28.61
<i>want to talk about</i>	Desire-Personal	28.14
<i>all over the world</i>	Other Referential	26.13
<i>when I was in</i>	Topic introduction/focus	26.13
<i>all of a sudden</i>	Topic elaboration/clarification	24.12
<i>I'm going to talk</i>	Intention/prediction	22.11
<i>so this is what</i>	Identification/focus	22.11
<i>going to tell you</i>	Intention/prediction	20.10
<i>I'm going to do</i>	Intention/prediction	20.10
<i>to figure out what</i>	Ability/effort	20.10
<i>to talk to you</i>	Intention/prediction	20.10
<i>why is it that</i>	Topic introduction/focus	20.10
<i>you can see that</i>	Ability/effort	20.10
<i>and this is how</i>	Identification/focus	18.09
<i>if you think about</i>	Topic introduction/focus	18.09

本研究では、TED Talks に特徴的であることが示された 48 個の 4 語の key lexical bundles (付表 2) の談話機能をこの基準に従って分類し、100 万語あたりの使用頻度をカウントした。図 1 は、これらの lexical bundles の頻度を談話機能別を示したものである。

表 4 Biber (2006) による lexical bundles の談話機能の分類

		Biber (2006) の例
I. Stance Expressions	A. Epistemic stance-Personal	<i>I don't know if, you know what I</i>
	B. Attitudinal/Modality stance	
	Desire-Personal	<i>if you want to, I just wanted to</i>
	Obligation/directive	<i>you have to do, I'd like you to, take a look at</i>
	Intention/prediction	<i>we're going to do, is going to be</i>
II. Discourse Organizers	Ability/effort-Personal	<i>to be able to, and then we can</i>
	Topic introduction/focus	<i>what do you think, you know if you</i>
	Topic elaboration/clarification	<i>at the same time, know what I mean</i>
	Intention/focus	<i>and this is a, one of the things</i>
III. Referential Expressions	Imprecision	<i>or something like that,</i>
	Quantity specification	<i>have a lot of, the rest of the</i>
	Multi-functional reference	<i>the end of the, at the end of</i>
IV. Special Functions	Politeness and inquiries	<i>thank you very much, do you have a</i>

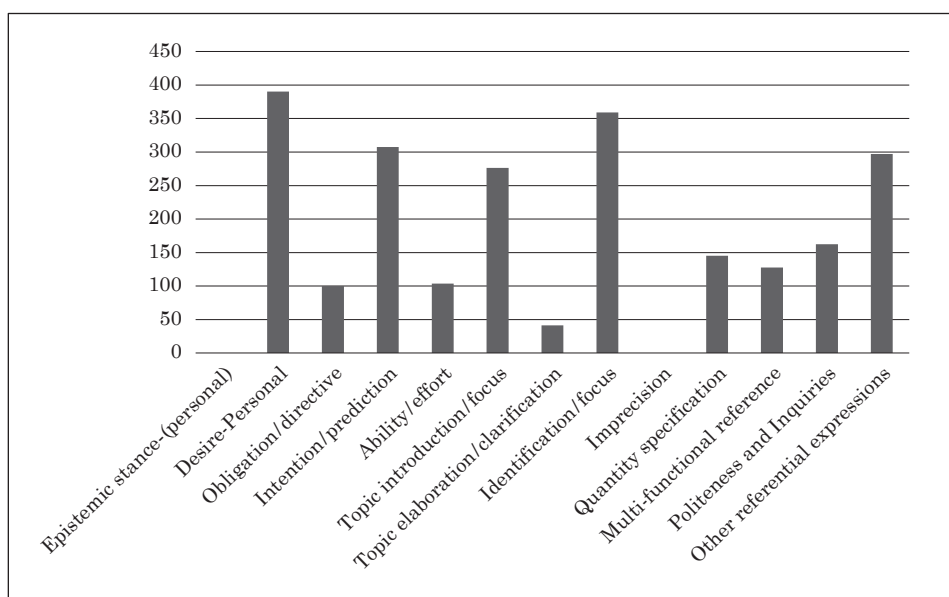


図 1 4 語の key lexical bundles の談話機能別頻度 (100 万語あたりの使用頻度)

これによれば, TED Talks に特徴的な lexical bundles には, 様々な談話機能を持つ lexical bundles が使用されていることが分かり, 特に Desire-Personal や Intention/prediction, Identification/focus の談話機能を持つものが多い傾向が見られるが, これは, TED Talks が自らの考えや研究, 活動などを一般聴衆に訴えるパブリックスピーチであることを反映していると考えられ, 参照コーパスの日常会話やビジネス会話などにおける使用頻度以上に TED

Talks ではこれらのフレーズが多く使用されることが示されている。また、このようなスピーチには聴衆の注目を効果的に集め、論理的に話すことが要求されるが、Topic introduction や elaboration, Identification/focus を果たす lexical bundles の使用が多いのはこのためであると推測される。一方で、動詞の *know* や *think* を含むような Epistemic stance を示す lexical bundles については、*know* を含むものは TED Talks コーパス全体でもそれほど出現頻度が多くなく、*think* を含むものについては比較的多く使用されていたものの、*I think + X* という形で複数の形の lexical bundles で使用されていたため、4 語の単体の lexical bundles としては対数尤度比検定で有意差が現れなかったものと考えられる。また、Imprecision の機能を持つ lexical bundles については、最初の n-gram 解析で抽出された 288 個の lexical bundles にも *something like that* 等を含むものは含まれておらず特徴フレーズの中にも存在しない。このような曖昧さを表すフレーズは、スピーチ内容に明確さや正確さが求められる TED Talks では避けられる傾向があるのがその原因ではないかと考えられるが、これについては今後の検証が必要である。

IV. まとめ

本研究では、3 つの research question を設定し、TED Talks で使用される 3 語及び 4 語の lexical bundles の分析を行ってその頻度表を作成すると共に、参照コーパスとの比較により、どのような lexical bundles が TED Talks に特徴的に使用されているかについての分析を行った。また、特徴的な lexical bundles がどのような談話機能を果たしているかを分類し、スピーカーがトークの談話を的確に構成して効果的に聴衆に訴えかけるために使用される lexical bundles の頻度が相対的に高い可能性を示した。英語教育においてフレーズレベルでの語彙指導を行う場合は、単にフレーズの意味を教えるだけで無く、それらが談話においてどのような機能を果たすかについても教えることが重要であることを本研究の結果は示しており、TED Talks を用いてリスニングやプレゼンテーションの指導を行う場合においては、これらの lexical bundles に注意させると共に、これらを正しく使用させたりその機能を理解させるような指導が重要である。

なお、本研究では代表的な TED Talks を集めたため、データサイズは比較的小規模なものであり、抽出した lexical bundles の数もまだ十分ではない。今後より大規模にデータを収集し、分析結果の信頼性を高める予定である。

<参考文献>

- Anthony, L. (2019). *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp>.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam, Philadelphia: John Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Coxhead, A., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *TESOLANZ Journal*, 20, 55-67.
- Liu, C. & Chen, H.H. (2020). Functional variation of lexical bundles in academic lectures and TED talks. *Register Studies*, 2(2), 176-208.
- Longshaw, R., & Blass, L. (2015). *21st century reading: Creative thinking with TED Talks*. Boston: National Geographic Learning/Cengage Learning.
- Shin, D. & Shin, K.Y. (2020). Retaliated with tariffs on: A corpus analysis of lexical bundles in TED Talks and BBC news on global business issues. *Modern English Education*, 21(2), 71-84.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins Publishing Company.
- Scotto di Carlo, G. (2014). The role of proximity in online popularizations: The case of TED talks. *Discourse Studies*, 16(5), 591-606.
- Wang, Y. (2012). An exploration of vocabulary knowledge in English short talks: A corpus-driven approach. *International Journal of English Linguistics*, 2(4), 3-43.
- Wingrove, P. (2017). How suitable are TED talks for academic listening? *Journal of English for Academic Purposes*, 30, 79-95.

付表 1

3-word lexical bundles (上位 100)

1	a lot of	35	we're going to	69	you very much
2	I'm going to	36	I was a	70	and if you
3	one of the	37	to tell you	71	and then I
4	I want to	38	we have a	72	at the end
5	this is a	39	some of the	73	to think about
6	this is the	40	you know what	74	all the time
7	and this is	41	you look at	75	are going to
8	a little bit	42	don't want to	76	people who are
9	going to be	43	there is a	77	to do this
10	when I was	44	to figure out	78	to talk about
11	in the world	45	we want to	79	want to be
12	you have to	46	you're going to	80	was going to
13	you can see	47	going to do	81	and you can
14	you want to	48	I have to	82	by the way
15	be able to	49	in the middle	83	I went to
16	it turns out	50	is going to	84	is that the
17	this is what	51	that we have	85	most of the
18	we have to	52	the rest of	86	that you can
19	the United States	53	percent of the	87	around the world
20	out of the	54	so this is	88	the first time
21	part of the	55	to do with	89	they're going to
22	we need to	56	all of the	90	to be the
23	of the world	57	the fact that	91	to make a
24	and I was	58	how do you	92	and a half
25	and I think	59	there was a	93	and I said
26	of the brain	60	you thank you	94	and she said
27	the end of	61	a couple of	95	and that is
28	to be a	62	and it was	96	if you look
29	I don't know	63	and so I	97	in the first
30	it was a	64	not going to	98	to look at
31	in front of	65	of the most	99	a kind of
32	I wanted to	66	thank you very	100	all of these
33	look at the	67	that we can		
34	the world and	68	to show you		

4-word lexical bundles (上位 100)

1	thank you very much	13	a little bit of	25	I'm going to show
2	thank you thank you	14	a lot of people	26	part of the brain
3	to be able to	15	in the middle of	27	the rest of the
4	at the end of	16	going to show you	28	when I was a
5	if you look at	17	I don't want to	29	when it comes to
6	in the United States	18	if you want to	30	you don't have to
7	one of the most	19	there's a lot of	31	and this is what
8	it turns out that	20	a lot of time	32	want to talk about
9	the end of the	21	for the first time	33	a few years ago
10	at the same time	22	how many of you	34	all over the world
11	I want to talk	23	I would like to	35	let me tell you
12	I want you to	24	and I'm going to	36	when I was in

37	you can see the	59	going to tell you	81	the top of the
38	all of a sudden	60	I can tell you	82	they're going to be
39	and one of the	61	I'm going to do	83	want to tell you
40	and this is a	62	not going to be	84	a little bit about
41	are going to be	63	the middle of the	85	and at the end
42	as you can see	64	to figure out what	86	and I wanted to
43	I was going to	65	to talk to you	87	are we going to
44	in front of the	66	why is it that	88	around the world and
45	in the first place	67	you can see that	89	for a long time
46	of the united states	68	you want to be	90	for those of you
47	one of the things	69	a couple of years	91	had a lot of
48	you look at the	70	a little bit more	92	I don't know if
49	I just want to	71	a lot of money	93	I want to tell
50	I'm going to talk	72	and this is how	94	I'm going to tell
51	in the world and	73	if you think about	95	in New York City
52	is going to be	74	in the case of	96	in the back of
53	is one of the	75	let me give you	97	just a little bit
54	so this is what	76	not supposed to be	98	most of the time
55	when you look at	77	on the other hand	99	of the things that
56	you thank you thank	78	one of the reasons	100	some of the most
57	and this is the	79	over and over again		
58	and you can see	80	so I want to		

付表 2 TED Talks における特徴フレーズ (4 語の lexical bundles)

1	it turns out that	17	why is it that	33	when you look at
2	I want to talk	18	you can see that	34	and this is a
3	going to show you	19	and this is how	35	if you look at
4	I'm going to show	20	if you think about	36	a few years ago
5	part of the brain	21	not supposed to be	37	and this is what
6	in the United States	22	over and over again	38	in the case of
7	want to talk about	23	so I want to	39	and you can see
8	all over the world	24	want to tell you	40	I don't want to
9	when I was in	25	thank you thank you	41	there's a lot of
10	all of a sudden	26	one of the most	42	how many of you
11	I'm going to talk	27	let me tell you	43	let me give you
12	so this is what	28	when I was a	44	when it comes to
13	going to tell you	29	of the United States	45	in the middle of
14	I'm going to do	30	in the world and	46	I just want to
15	to figure out what	31	I want you to	47	you want to be
16	to talk to you	32	and I'm going to	48	the top of the