

チュートリアル

TCSE を用いた TED Talks の全文検索と英語学習への応用

長谷部 陽一郎

1 はじめに

ここでは本チュートリアルの目的を示し、TCSE と TED Talks に関する基本的事項を確認します。

1.1 本チュートリアルの目的

- TCSE を用いた TED Talks 全文検索の基本的な方法を知る
- TCSE の様々な機能について知る
- TCSE を英語学習に応用する方法について考える

1.2 TCSE とは何か

TED Corpus Search Engine (TCSE) は TED (<https://ted.com>) が公開している 2,700 件以上のプレゼンテーションの英語トランスクリプトと多言語翻訳テキストを解析してデータベースに格納し、全文検索を可能にしたシステムです。

2014 年 11 月に最初のバージョンを公開して以来、基本的な検索に加えて、言語研究や英語学習に役立つ様々な機能を実装してきました。TCSE は下記の URL でアクセスできます。

<https://yohasebe.com/tcse>

演習 1

上記 URL で TCSE にアクセスして、テキストボックスに好きな英単語を入力し、SEARCH ボタンをクリックしてみましょう。

TIPS RESET ボタンをクリックすると新たな検索を行うことができます。

TCSE の各種機能（一部の実験的機能を除く）は Windows, Android, MacOS, iOS の Google

Chrome ブラウザで動作確認をしています。

1.3 TED Talks について

TED is a nonprofit devoted to spreading ideas, usually in the form of short, powerful talks (18 minutes or less). TED began in 1984 as a conference where Technology, Entertainment and Design converged, and today covers almost all topics —from science to business to global issues —in more than 100 languages. Meanwhile, independently run TEDx events help share ideas in communities around the world.

<https://www.ted.com/about/our-organization>

TED から提供されているデータ

- 2,700 以上のプレゼンテーションの動画・音声ファイル
- TED による書き起こしデータ（英語トランスクリプト）
- ボランティアによる多言語翻訳テキスト（ピア・レビュー済）
- 各 Talk の視聴数や評価など様々な関連情報

TED データの利用規定

- データは Creative Commons ライセンス（BY-NC-ND）で再利用・再配布可能
- 出典を明記し（BY）、非商用（NC）で、動画の改変を行わない（ND）こと

<https://www.ted.com/about/our-organization/our-policies-terms>

演習 2

次のページにアクセスして、TED Talks のトランスクリプトの多言語翻訳がどのような仕組みで行われているか調べてみましょう。 <https://www.ted.com/participate/translate>

TED API について

以前は開発者が TED データを用いたアプリケーションを作成するための公式 API が公開されていました。TCSE も公式 API を利用する形で開発が進められてきました。現在 API の新規利用申請はできません。

1.4 TCSE の基本統計情報

TED Talks の数はコンスタントに増え続けています。TCSE のデータベースは月に 1 度くらいの頻度で更新を行っています。

表1 TCSE の基本統計情報 (2018 年 6 月 30 日現在)

種別	数
利用可能な TED Talks	2,721
翻訳テキスト言語	28
セグメント (segments)	728,654
拡張セグメント (expanded segments)	326,964
語彙的要素 (elements) ¹	6,497,466
語彙項目 (lexical items) ²	86,176

¹ 概ね「延べ語数」に対応 (ただし、記号類などもカウント)

² 概ね「異なり語数」に対応

2 TCSE の基本的な使い方

ここでは TCSE を使った全文検索の基本的な方法を示します。

The screenshot shows the TCSE search interface. At the top, there is a blue button that says "Input text and press SEARCH". Below this is a search input field and two buttons: "SEARCH" and "RESET". Underneath the input field, there are several settings:

- "Translation" is set to "Japanese" with a dropdown arrow.
- "Search Target" has radio buttons for "English" (selected) and "Translation".
- There are checkboxes for "Advanced Search (English only)", "Search Talk Info (speaker/title/description)", "Include English only talks" (checked), and "Use Expanded Segments".
- There are checkboxes for "Play video 10 seconds earlier" and "Stop automatically".
- "Play speed" is set to "x 1.00" with a dropdown arrow.
- "Video source" has radio buttons for "TED" and "YouTube" (selected).
- "TED video quality" has radio buttons for "Normal" (selected) and "High".

At the bottom, there are two buttons: "List all available talks" and "I'm feeling lucky".

図1 TCSE の基本操作パネル

2.1 英語トランスクリプトと翻訳テキストの検索

The screenshot shows a search interface with the following elements:

- ①: A large text input box for the search query.
- ②: A blue button labeled "SEARCH".
- ③: A button labeled "RESET".
- ④: A dropdown menu labeled "Translation" with "Japanese" selected.
- ⑤: Radio buttons for "Search Target", with "English" selected and "Translation" unselected.

図2 検索用テキストボックスとトランスクリプト／翻訳テキスト設定

図2の説明

- ① 検索用テキストボックス
- ② 検索実行ボタン
- ③ リセットボタン (新たな検索を行う際に必ずクリックする必要があります)
- ④ 翻訳言語セレクタ (OS およびブラウザの言語設定に従います)
- ⑤ 検索対象セレクタ (英語トランスクリプトまたは翻訳言語を指定します)

演習 3

次の文字列を検索してみましょう。アスタリスク (*) でワイルドカード検索ができます。

- (a) expect
- (b) no matter wh
- (c) do you * ?

TIPS 通常検索では検索文字列は単語の境界に関係なくマッチします。文字列の先頭や末尾にスペースを入れると単語の境界を考慮した検索が可能です (3節で示す **Advanced Search** を用いるとさらに高度な検索ができます)。

演習 4

Search Target を **Translation** にセットして日本語の語句を検索してみましょう。時事的なキーワードを検索するのもよいですし、英語と日本語で一対一対応しない機能表現を検索してみるのも面白いでしょう。

- (a) 自動運転, 人工知能, 仮想現実, 格差
- (b) もしかして, どうしても, 何となく

TCSE で利用可能な翻訳言語

現在、TCSE では 28 の翻訳言語が利用可能です。現時点で翻訳済みの Talk 数が 1,000 以上ある言語を選定しています。各言語で TED Talks がどれだけ翻訳されているかは次の URL を参照してください。 <https://www.ted.com/participate/translate/our-languages>

演習 5

Translation セレクタで日本語以外の言語（中国語、韓国語、スペイン語、ロシア語、etc.）を設定した上で、英語トランスクリプトを検索してみましょう。さらに可能であれば、それらの言語による検索も試してみましょう。

TIPS あるセグメントが複数の言語でどのように訳されているかを確認するには、次節の図3⑥をクリックします。利用可能なすべての言語による訳がリストアップされます。（ただし、TED Talks の翻訳はセグメント・レベルで一対一対応しているとは限りません。自然な訳とするために、前後が入れ替わっていたりすることが多くあります。）

2.2 検索結果の表示

#	ID	Line	Time				
Total : 51 hits in 41 / 2431 talks [0.396 seconds] Help Prev 100 Next 100							
①							
1	2852	25	01:44	☰ ▶ 🔗	of the University of Tokyo,		
	②	[0.11]	[13:24]	③ ④ ⑤	⑥		
2	2852	38	02:35	☰ ▶ 🔗	To enter Todai, the University of Tokyo,		
		[0.17]	[13:24]				
3	2792	71	04:42	☰ ▶ 🔗	in an abandoned parking lot outside Tokyo,	傘を持った何百人という人に 振りを付け それを 700メートル上空から ドローンで撮影しました	
		[0.21]	[17:18]				
4	2591	171	09:40	☰ ▶ 🔗	the lights in London, the lights in Cairo, the lights in Tokyo,	ロンドンの光 カイロの光 東京の光	
		[0.68]	[13:50]				
5	2591	184	10:23	☰ ▶ 🔗	you would see that Tokyo has actually become darker,	東京が暗くなったことが分かるでしょう	
		[0.73]	[13:50]				

図3 検索結果の例

図3の説明

- ① 検索ヒット数の表示 (赤字はそれぞれヒットした事例数, ヒットした Talk 数, 総 Talk 数)
- ② Talk ID (クリックすると当該の Talk とスピーカーの情報をポップアップ表示します)
- ③ パラグラフ形式での全文表示ウィンドウをポップアップ
- ④ ピンポイント動画再生 (詳しくは 2.4 節で解説)
- ⑤ 動画再生用 URL 生成 (当該の動画再生画面に後からアクセスするのに便利)
- ⑥ 検索文字列を含むセグメントの表示 (クリックすると自動解析された統語情報を表示)

演習 6

好きな文字列を検索し, 表示されたセグメントのいくつかの動画を再生してみましょう.

TIPS ポップアップウィンドウ上で画面をクリックすると再生が始まります. 同じ Talk 内の別のセグメントを再生したいときには, いったん一時停止して, 再生したいセグメントのトランスクリプトをクリックしましょう.

演習 7

TED Talks のスピーカーは英語のネイティブスピーカーばかりではありません. Talk ID をクリックして, Talk の概要とスピーカー情報を確認し, 使われている英語がどのような言語的背景を持つものかを推測してみましょう.

2.3 様々な検索オプション

①
②

Advanced Search (English only)
 Search Talk Info (speaker/title/description)

Include English only talks
 Use Expanded Segments

③
④

図4 検索詳細設定

図4の説明

- ① 高度な検索を行う (詳しくは 3 節で解説します)
- ② Talk の本文ではなくスピーカー名と概要テキストを検索します (英語 ⇄ 翻訳)
- ③ 指定した言語による翻訳が存在しない Talk も検索対象とする
- ④ 検索対象を少なくとも 1 つのセンテンスの全体を含む形 (拡張セグメント) に設定

演習 8

図 4 の②をクリックして Talk 情報検索モードにした上で、スピーカー名または概要テキストに次の文字列を含む Talk を検索してみましょう。他の適当なキーワードについても試してみましょう。

- (a) Japan
- (b) blockchain
- (c) Elon Musk

2.4 動画再生関連のオプション

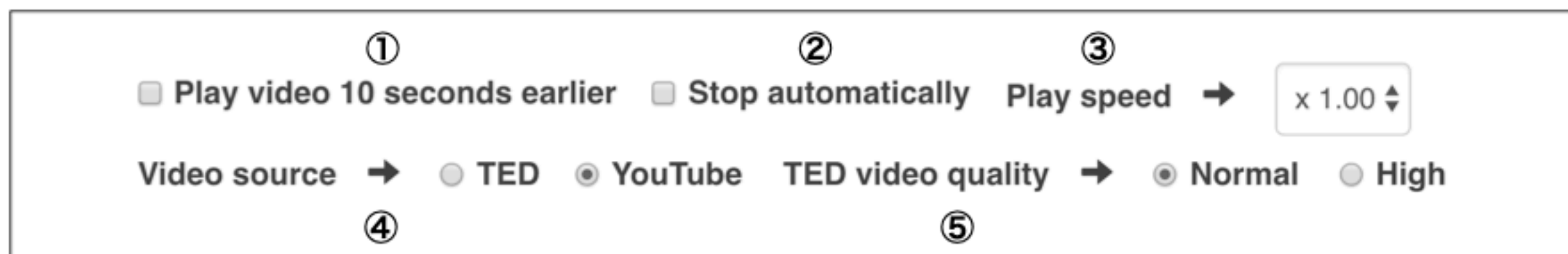


図5 動画再生関連のオプション

図5の説明

- ① 当該箇所の 10 秒前からピンポイント再生します
- ② 当該箇所を再生後に自動的に停止します
- ③ 動画再生速度の設定
- ④ 動画ソースの選択（デフォルトは YouTube, 将来的に TED 公式のオプションは外す予定）
- ⑤ 動画ソースが TED 公式の場合の動画品質の変更

演習 9

適当な文字列で検索を行い、表示されたセグメントをピンポイント再生してみましょう。

Talk によっては音声とトランスクリプトの表示が微妙にずれている場合があります。いくつかの動画を再生して、そのような Talk を見つけたらメモしておきましょう。後ほど、音声とトランスクリプトを手動で同期させる方法を示します。

2.5 Talk の選択を補助する機能

特定の表現を探しているのではなく、単純に英語学習の素材となる Talk を選びたいときに役立ついくつかの機能を紹介します。

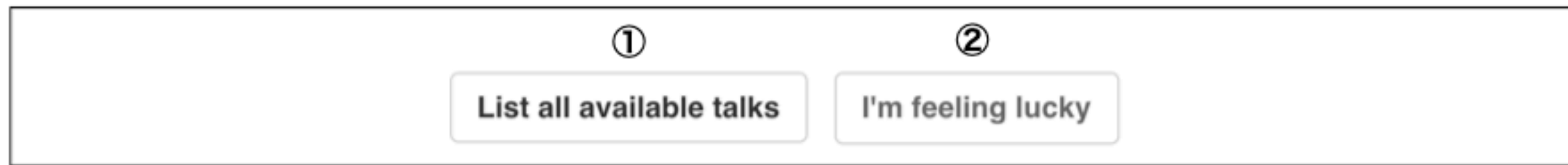


図6 Talk 選択補助関連の機能

図6の説明

- ① TCSE で利用可能なすべての Talk を一覧表示します
- ② すべての Talk の中からランダムに選んだ Talk を再生します

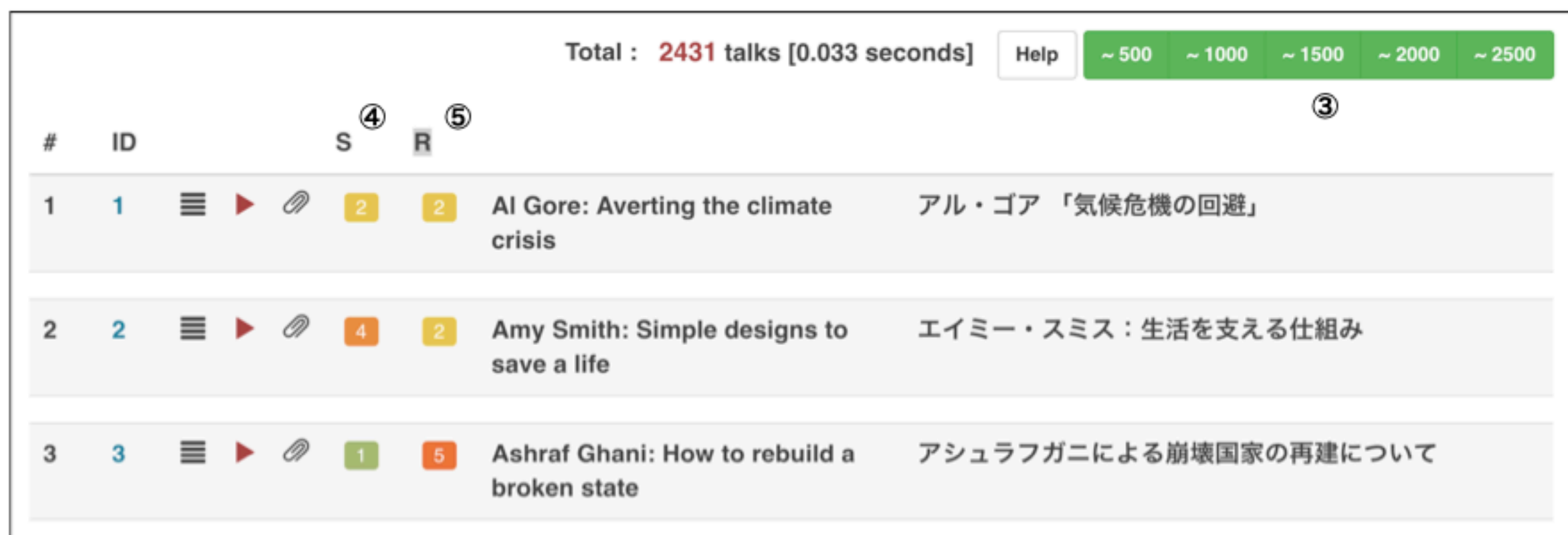


図7 全 Talk のリストアップ画面 (抜粋)

図7の説明

- ③ 全 Talk を 500 件ごとにリストアップして表示します
- ④ 発話速度ランク：1分あたりの平均語数 (wpm) を元に、全 Talk を 1 (最遅) から 5 (最速) までの 5 ランクに分類しています
- ⑤ 難易度ランク：Flesch-Kincaid grade level を算出し、全 Talk を 1 (最易) から 5 (最難) までの 5 ランクに分類しています

Flesch-Kincaid grade level

$$0.39 \left(\frac{\text{total words}}{\text{total sentences}} \right) + 11.8 \left(\frac{\text{total syllables}}{\text{total words}} \right) - 15.59$$

https://en.wikipedia.org/wiki/FleschKincaid_readability_tests

演習 10

TCSE の全 Talk リストアップ画面で、発話速度と難易度のランクが共に 3 である Talk, すなわち、発話速度と難易度に関して最も平均的とみられる Talk を探し、それらがどのようなものか確認してみましょう。逆に 1 や 5 など極端な数字を示している Talk もいくつか見てみましょう。

TIPS TED Talks のトランスクリプトには、発話内容以外の情報も含まれているため、発話速度と難易度の数値は必ずしも正確なものではありません。楽器演奏の部分や、「笑い」などオーディエンスの反応に関する部分などは除外して計算していますが、あくまで「目安」と考えてください。

3 高度な検索

ここでは、表層形だけでなく、レンマ、品詞情報、ワイルドカードなどを用いたより高度な検索を行う方法を示します。

3.1 検索式を用いた検索

操作パネルで **Advanced Search** を選択すると、英語トランスクリプトに対する高度な検索が可能になります。下記の検索式は基本的にそれぞれが「1 語」を構成することに注意してください。2 語以上の句を検索する場合にはスペースを検索式の間に入挿入する必要があります。

利用可能な演算子と利用例

- ① **レンマ検索** [LEMMA]
[excite] ⇒ excite, excites, exciting, excited
- ② **品詞検索** {POS}
{vb} ⇒ 動詞
{nn} ⇒ 名詞
注：形容詞 = {jj}, 副詞 = {rb}, 代名詞 = {pr}, 前置詞 = {in}
その他の品詞コード：<https://yohasebe.com/tcse/pos>
- ③ **OR 検索** A|B
[read] {dt} [news|article]
⇒ I'm reading the news at six
⇒ they read these articles

④ **表層形 + 品詞検索** SURFACE{POS}

surfing{nn} ⇒ it's like surfing

surfing{vb} ⇒ they're surfing the web on their phones

注：括弧類なしの文字列は表層形とみなされます

表層形文字列と {} の間には空白を入れしないでください

⑤ **レンマ + 品詞検索** [LEMMA]{POS}

[help]{nn} ⇒ an aunt offered financial help

[help]{vb} ⇒ to help people become their best selves

注：[] と {} の間には空白を入れしないでください

⑥ **ワイルドカード** *

to * surprise ⇒ to our surprise, to many people's surprise

注：* は 1 語以上の幅の文字列に対応します

ちょうど 1 語の幅を指定したいときには下の否定一致検索を参考にしてください

⑦ **セグメント頭** ^

^ having {vb}

⇒ Having started the process ...

⇒ Having said that ...

注：Expanded Segments を指定している場合には全体の先頭を意味します

⑧ **前方一致検索** PARTIAL+

over+

⇒ overlook, overwhelmed, overview, etc.

⑨ **後方一致検索** +PARTIAL

[get] +ed

⇒ I get asked sometimes

⇒ let me tell you how we got started

⑩ **前方一致 + 品詞検索** PARTIAL+{POS}

socio+{jj}

⇒ sociological, socioeconomic, sociopolitical, etc.

注：前方一致文字列と {} の間には空白を入れしないでください

⑪ 否定一致検索 -PARTIAL

[get] -rid of

⇒ get outside of, get ahead of, got tired of, etc. (get rid of は含まれない)

注：否定一致検索では一致する文字列を含む要素が結果から除外されます。

[get] -_ of

⇒ get x of のパターン全体

注：_ を含む単語が事実上存在しないことを利用し，-_ で1語の文字列を指定できます。

演習 11

TCSE の検索機能を駆使して次のような熟語や構文の事例を採取してみましょう。

- (a) [look forward to] の様々なバリエーション
- (b) [should/could/might/would/must + have + 動詞] のパターン
- (c) [nor do/does/did] のパターン
- (d) [get + 代名詞 + -ed] のパターン (代名詞={pr})
- (e) セグメントの最初 (≈ 文頭) に生起する [without,]

TIPS Advanced Search ではピリオドやコンマなどの記号も1語として扱うため、前(後)にスペースを入れる必要があります。

注意！

TCSE では統語解析プログラムを用いてコーパス・データの品詞解析を行っています。解析結果は完璧ではありません。数値的な正確性が求められる調査を行う場合には、表層形やレンマを検索キーとして用いることをおすすめします。

4 TCSE の英語学習への応用

ここでは TED Talks を英語学習に用いるのに役立つ TCSE の機能と利用例を見ていきます。

4.1 動画を利用した事例の提示

TCSE では検索結果に対応する動画をピンポイントで再生することができます。話し言葉ならではの音声的なニュアンスや、表現に付随する非言語的な要素を確認しながら、事例をみていくことができます。TCSE で動画をコントロールするには画面右下の操作パネル、もしくはショートカット・キーを用います。Talk 内の別の箇所を再生するときは、いったん一時停止し、再生したい箇所のトランスクリプト文字列をクリックします。

TCSE の高度な検索機能と事例の生起箇所をピンポイントで再生する機能を用いると、辞書的な定義や説明だけでは習得が難しい談話標識の使い方などを学ぶのに役立つ素材を採取することができます。

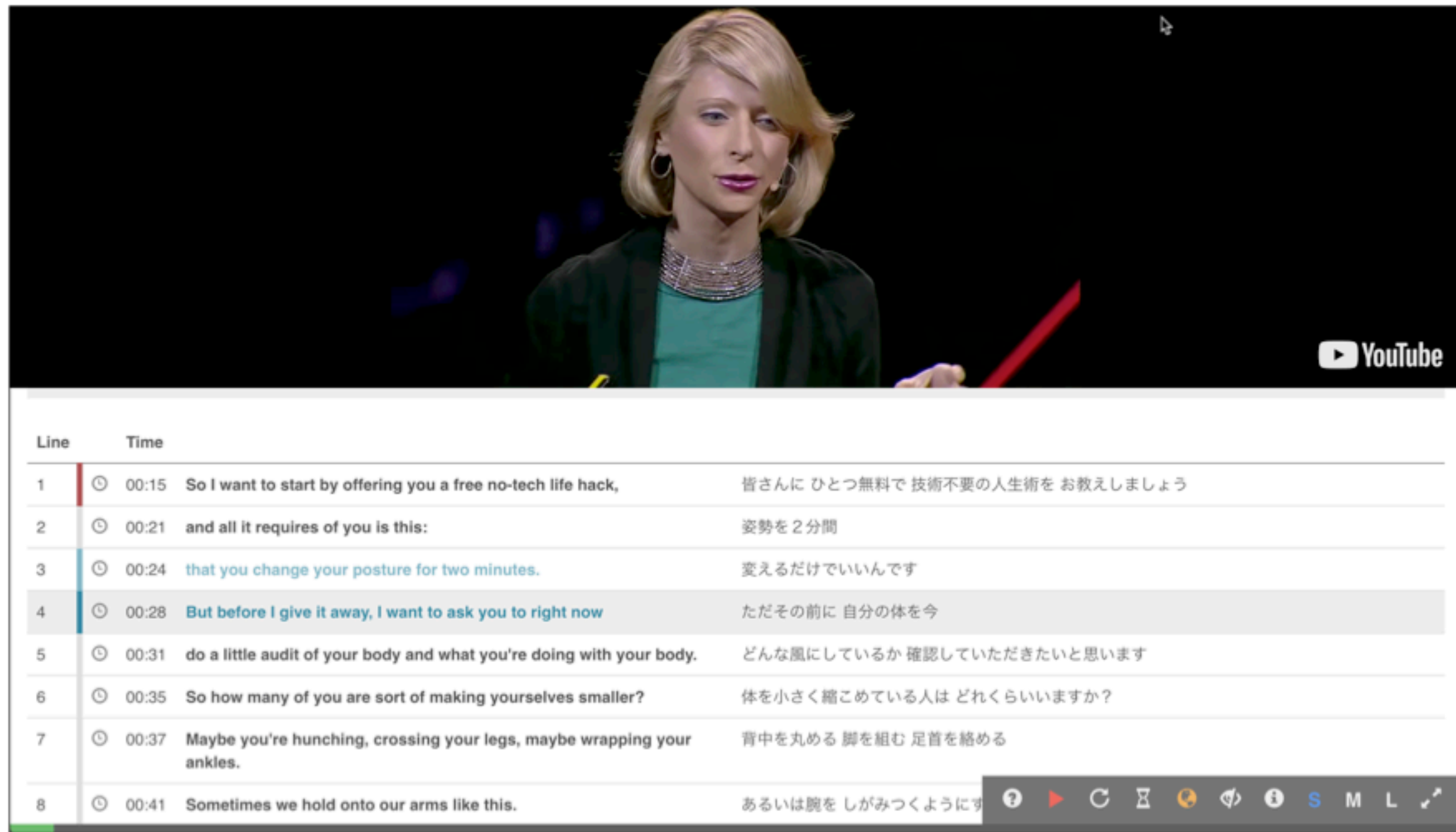


図8 TCSE の動画再生画面

演習 12

英語の話し言葉でよく使われるイディオム、談話標識、構文をいくつか検索して、発話部分の動画を視聴してみましょう。通常検索モードで次のような文字列を入力してみてください。(ワイルドカードの * は通常検索モードでも使用可能です。)

- (a) 文末における関係代名詞の非制限適用法 ⇒ , which is * .
- (b) 付加疑問文の [isn't it?] ⇒ isn't it?
- (c) 「いわば」の [if you will] ⇒ , if you will

また、Advanced Search モードで次のような文字列を入力してみましょう。このモードではカンマなどの記号の前にスペースが必要なことに注意してください。

- (d) 注意喚起の [Look,] ⇒ ^ look ,
- (e) 副詞的な [as X as] ⇒ ^ as -_ as
- (f) [not only ... but] の構文 ⇒ [not] only * but

演習 13

検索結果画面に表示される URL 生成アイコン (図3 ⑤) をクリックして、動画をピンポイント再生するための URL を生成してみましょう。また、実際にその URL にアクセスして動画を再生してみましょう。

4.2 TCSE を用いたリスニング/スピーキング学習

TCSE には TED Talks を英語のリスニングやスピーキングの学習に役立てるためのいくつかの機能があります。

4.2.1 TCSE の動画再生操作パネル

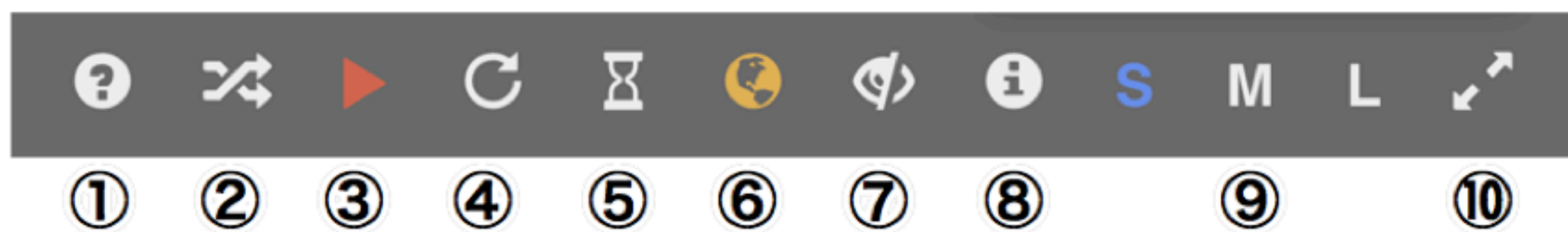


図9 動画再生操作パネル

図9の説明

- ① ヘルプ：各キーの説明とショートカット・キーを表示 (H)
- ② ランダム選択：ランダムに選択されたトークを呼び出す
- ③ 再生・停止ボタン (SPACE)
- ④ 直前 (3 秒前) に巻き戻して再生 (R)
- ⑤ 自動ポーズ機能のオン・オフ (A)
- ⑥ 翻訳テキストの表示・非表示 (T)
- ⑦ トランスクリプト全体の表示・非表示 (/)
- ⑧ 一時停止して Talk 情報を表示 (I)
- ⑨ フォント小・中・大 (S/M/L)
- ⑩ 全画面表示 (ESC)

* 括弧内はショートカット・キー (実際の打鍵は小文字で)

演習 14

PC 上で I'm feeling lucky ボタンをクリックしてランダムな動画を再生し、キーボードショートカットで「再生・停止」「翻訳テキストの表示・非表示」「トランスクリプト全体の表示・非表示」「全画面表示オン・オフ」などを試してみましょう。

TIPS 全画面モードはスマートフォンやタブレット端末でも使えますが、パケット通信データを消費するので WIFI 環境以外では注意してください。スマートフォンで全画面表示する際には画面をランドスケープ・モード（横向き）にしてください。

4.2.2 音声とトランスクリプトの同期調整方法

55	🕒 02:52	And this is what you might see:	
56	🕒 02:55	sunlight glinting off of water, you might hear gulls flying above,	
57	🕒 02:59	you might even feel the wind on your cheeks or smell seawater,	
58	🕒 03:02	with advanced hardware.	①
59	🕒 03:04	Now, if you were to spend a lot of time here,	
60	🕒 03:06	you might want a dashboard	②
61	🕒 ←	where you could do things like make a phone call,	③

図10 同期の調整

音声とトランスクリプトの同期は次の手順で行います。ここでは図10の①がハイライトされているときに音声とのズレに気付いたとします。

同期調整の手順

- 聞こえてくる音声と、ハイライトされているセグメントがずれている … ①
- 少し先のセグメントで同期させるべく、一時停止して方針を検討する
- ③のポイントで一致させることに決めたら一時停止を解除
- 耳をたよりに、③の内容が発話される直前で再び一時停止する … ②
- セグメントの左側にある時計のアイコンをクリック … ③
- ③の再生が開始されると共に正しくハイライトが行われ、調整完了

演習 15

先の演習 9 でメモした、音声とトランスクリプトの同期が完全でない Talk を再び呼び出して、同期の微調整を試みてみましょう。

4.2.3 ポーズ・アンド・チェック機能

英語のリスニング能力の訓練を行う際、画面上に字幕があったり、印刷されたテキストが手元にあったりすると、知らず知らずのうちにそれらに依存して、耳からの入力に対し十分な意識を向けられないことがあります。TCSE の「ポーズ・アンド・チェック機能」はこの問題を解消するための機能です。

ポーズ・アンド・チェック機能を用いると、聞き取れない箇所が生じたときだけ一時停止し、画面に表示された直前のセグメントの英語トランスクリプト（+翻訳テキスト）を確認することができます。



図11 ポーズ・アンド・チェック機能

図11の説明

- ① まずは全画面ボタンをクリックして全画面モードに
- ② 全画面表示での再生中はトランスクリプトが一切表示されません
- ③ 聞き取れない箇所があれば一時停止します
- ④ 画面に現在および直前のセグメントのトランスクリプト（+翻訳テキスト）が表示されます
- ⑤ 一時停止を解除すると再び表示が消え、再生が始まります

演習 16

ポーズ・アンド・チェック機能を使いながら TED Talks を視聴してみましょう。一通り感覚をつかんだら、**Use Expanded segments** にチェックを入れた状態でも試してみましょう。

5 TCSE の実験的機能

ここでは、現在試験的に公開しているいくつかの実験的機能を紹介します。実験的機能は今後仕様が大幅に変更されたり、一部機能が停止されたりすることがあります。

5.1 自動ポーズ機能

TCSE の自動ポーズ機能をオンにすると、セグメントが発話されるたびに動画が一時停止します。一時停止の解除は操作パネルまたはショートカットキー（SPACE）を用います。動画自体をクリックしても一時停止を解除できません。

自動ポーズ機能は英語のディクテーションやリピーティングの練習に用いることを想定しています。図12に示すように、目的に合わせて、テキストを表示させたり、完全に非表示にしたりして利用することができます。

自動ポーズ機能は全画面モードでも利用できます。その場合には、一時停止しているときだけテキストが表示されます。

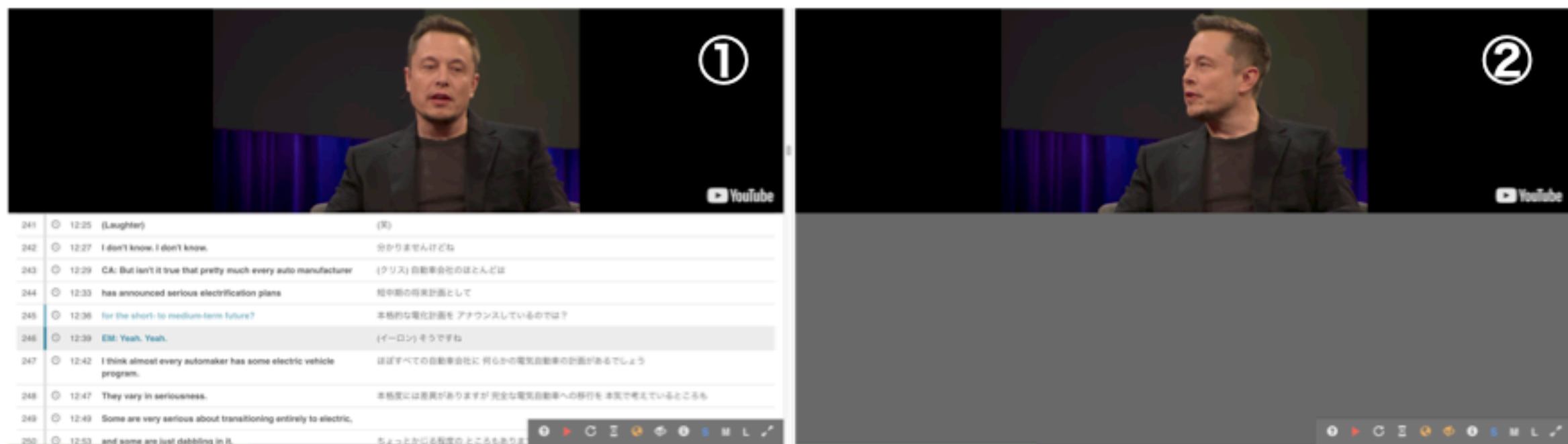


図12 テキストの表示

図12の説明

- ① トランスクリプトを常に表示：内容と音声をじっくり確認するのに適しています。
- ② トランスクリプトを常に非表示：リピーティングの練習などに適しています。

演習 17

自動ポーズ機能が役立つするのは、トランスクリプトが正確に同期する Talk に限られます。いくつかの Talk を再生して、うまく動作するものを探してみましょう。

また、4.2.2節で示した方法を使い、できるだけ同期を調整してから自動ポーズ機能をオンにして試してみましょう。

5.2 n-gram 検索機能

n-gram 検索モードでは、TCSE のデータベースに収められているデータの 2-gram, 3-gram, 4-gram のデータを検索・閲覧することができます。n-gram 検索モードに移行するには、画面上部のモード・セレクトで n-gram を選択します。

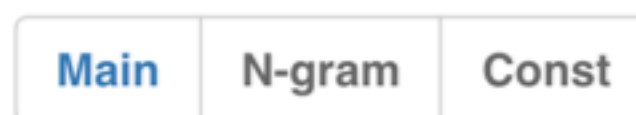


図13 モード・セレクト

ngram 検索できるのは単語の表層形のみです。表層形なので、exciting と excited は区別されます。また、get excited のようなフレーズは検索できません。

2-gram #2					3-gram ALL							
	Word1	Word2	Target POS	Freq	Num of Talks		Word1	Word2	Word3	Target POS	Freq	Num of Talks
1	the	opportunity	{nn}	223	184	1	i	was	surprised	{vb}	22	22
2	an	opportunity	{nn}	167	130	2	be	surprised	to	{vb}	20	19
3	of	opportunity	{nn}	41	33	3	might	be	surprised	{vb}	12	11
4	this	opportunity	{nn}	30	29	4	surprised	to	know	{vb}	10	10
5	and	opportunity	{nn}	15	13	5	be	surprised	if	{vb}	9	9
6	equal	opportunity	{nn}	11	10	6	n't	be	surprised	{vb}	9	9
7	great	opportunity	{nn}	10	10	7	be	surprised	by	{vb}	8	8
8	that	opportunity	{nn}	10	10	8	was	surprised	to	{vb}	8	8
9	unique	opportunity	{nn}	9	9	9	surprised	to	learn	{vb}	7	7
10	incredible	opportunity	{nn}	9	8	10	to	be	surprised	{vb}	7	6
11	every	opportunity	{nn}	7	7	11	surprised	to	see	{vb}	7	6
12	amazing	opportunity	{nn}	7	7	12	surprised	to	find	{vb}	6	6
13	that	opportunity	{nn}	7	7	13	may	be	surprised	{vb}	6	5
14	tremendous	opportunity	{nn}	7	6	14	was	surprised	that	{vb}	5	5
15	real	opportunity	{nn}	6	6	15	really	surprised	me	{vb}	5	5

図14 左: opportunity [2-gram RIGHT] 右: surprised [3-gram ALL]

n-gram 検索を行った後にリストアップされる n-gram 文字列をクリックすると、Advanced Search モードに移行し、当該の文字列を検索した結果が表示されます。

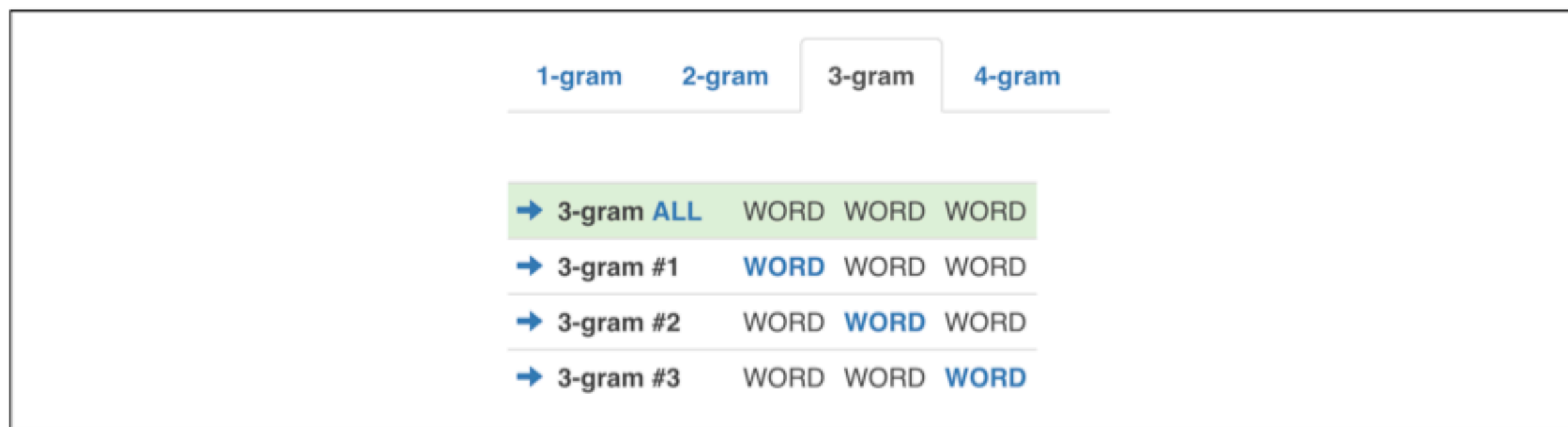


図15 n-gram 検索結果の表示選択

5.3 構文検索機能

構文検索モードでは、学習者にとって重要と考えられる英語の構文 (constructions) やイディオムを検索し、それらを TCSE で検索するためのリンクと例文 (作例) を得ることができます。構文検索モードに移行するには、画面上部のモード・セレクトで **const** を選択します。

現在、約 1,000 の構文と熟語を登録しています。構文および熟語のリストと例文はオリジナルのもので (選定は多分に恣意的)、TED と同様に Creative Commons ライセンス (BY-NC-ND) としています。

List of Constructions	
1	<i>I wish V</i>
2	<i>I be afraid</i>
3	<i>a bunch of</i>
4	<i>a great deal</i>
5	<i>a handful of</i>
6	<i>a host of</i>
7	<i>a variety of</i>
8	<i>about to V</i>
9	<i>about / high time V</i>
10	<i>above all</i>
11	<i>absorbed in</i>
12	<i>absorbed in</i>
13	<i>abundant in</i>
14	<i>according to X</i>
15	<i>accustomed to V-ing</i>
581	<i>what to V</i>
582	<i>what X call</i>
583	<i>what X be</i>
584	<i>what be called</i>
585	<i>whatever X</i>
586	<i>whatever happen</i>
587	<i>when it comes to X</i>
588	<i>when V</i>
589	<i>when / while V-ing</i>
590	<i>whenever V</i>
591	<i>where to V</i>
592	<i>whether or</i>
593	<i>whether to V</i>
594	<i>whether to V</i>
595	<i>whichever X</i>

図16 構文リスト表示

構文検索の結果画面上部には、**Advanced Search** モードでその構文を検索するための文字列が示されます。その下に示される例文は合成音声による読み上げ再生が可能です。



図17 事例検索用リンクと例文

6 おわりに

6.1 まとめ

本チュートリアルでは次の内容を扱いました。

- TCSE を用いた TED Talks 全文検索の基本的な方法
- TCSE の様々な機能
- TCSE を英語学習に応用する方法

謝辞

本研究の一部は科学研究費 (25870898, 18K00670) の補助を受けて行われました。

参考文献

- [1] Hasebe, Yoichiro. (2015) Design and implementation of an online corpus of presentation transcripts of TED Talks. *Procedia: Social and Behavioral Sciences*, 198(24). 74–182.
- [2] 長谷部陽一郎. (2018) 「TED Crpus Search Engine: TED Talks を研究と教育に活用するためのプラットフォーム」『英語コーパス研究』 25, 159-171.