

2016 年度関東支部研究支援プログラム中間報告
日本人英語学習者のための
TED プレゼンレベル計測方法の開発：
リーダビリティ，語彙レベル，発話速度，構音速度をもとに

田 淵 龍 二 （ミント音声教育研究所）
山 口 高 領 （早稲田大学）

1. はじめに

同時代の専門家による質のよいプレゼンが，TED Talks として公開されている。主な言語は英語で，多言語による字幕も付いている。ICT 環境の進展とともに，これらの動画を英語学習に活用する動きは次第に高まってきたようだ。そこで問題となるのが教材としての動画の選択である。TED Talks のページでは，動画の長さや分野での選別が可能になっている。しかし，いざ再生してみると，かなりの早口であったり，なじみのない語彙が多かったりと，担当クラスにはそぐわないこともある。特に初級者クラスの場合，選択の幅が狭くなる。そこで筆者らは，学習者レベルにあった TED 動画を手軽に選別できる仕組みの開発を始めた。

2. 方法

TED 動画を学習のしやすさで選別する指標としては，数値化が可能なリーダビリティ，語彙レベル，発話速度，構音速度の 4 つとした。リーダビリティは作動記憶の容量（時間と個数）制限にかかわらず，語彙レベルは主に音韻符号化速度と意味理解にかかわらず，発話速度は主に作動記憶の処理速度にかかわらず，構音速度は主に音声の分節化にかかわる指標と考えられる。ちなみに，発話速度はポーズも含めた時間で計算し，構音速度はポーズを含めないで計算したものである。

リーダビリティはテキストの文としての読みやすさを表す指標で，英語母語話者向けのフレッシュ・キンケイド学年指標が有名であるが，今回は日本人英語学習者向けの適応学年指標である MGJP を使った。MGJP は式①で表される。

$$MGJP = 0.07496 * (3 * SyPP + 2 * CPP) + 7.926 * LOG(PPC) + 4.618 \dots \textcircled{1}$$

where MGJP: an estimated Reading Grade Level for Japanese students

SyPP: Syllables Per Phrase / 1 フレーズに含まれる平均音節数

CPP: Consonants Per Phrase / 1 フレーズに含まれる平均子音数

PPC: Phrases Per Clause / 1 節（文）に含まれる平均フレーズ数

LOG: Common Logarithm / 常用対数

語彙レベルは単語の親密度を表す指標で，出現頻度を元にした NBC や JACET8000 が有名であるが，ここでは日本人英語学習者向けの適応学年指標を使った。この適応学年指標 VGL は式②で表される。

$$VGL = 30.371 * H + 8.7914 \dots \textcircled{2}$$

where VGL: an estimated Vocabulary Grade Level for Japanese students

H: a percentage of senior high school level words

発話速度はポーズも含めた話す速度で，1 分あたりの単語数で計測し，構音速度はポーズを含めない速度で，ひと息の連続音声（breath group）の 1 秒あたりの

単語数で計測した。発話速度と構音速度の適応学年指標を求める式を③④に示す。4つの公式のうち、②VGL, ③SPGL, ④ARGLは有効性を検証中である。

$$\text{SPGL} = 0.1815 \text{ wpm} - 13.416 \dots \textcircled{3}$$

$$\text{ARGL} = 8.4073 \text{ wps} - 9.7197 \dots \textcircled{4}$$

where SPGL: an estimated Speech Rate Grade Level for Japanese students

wpm: words per minute

ARGL: an estimated Articulation Rate Grade Level for Japanese students

wps: words per second

TED Talks から 2,143 本の英語字幕ファイル (字幕約 60 万本, 約 450 万単語) を取得し, 4つの指標の相乗平均を総合的な適応学年 (GGL, Grand Grade Level) とした。

3. 結果

解析結果の基本統計量, 平均 13.2, 標準偏差 1.9 であった。度数分布を表 1 と図 1 に示す。概ね高校 3 年から大学 3 年までに 4 分の 3 が集まっていた。本数が減っているのは音楽などのプレゼンを省いたからである。

表 1. TED Talks の MGJP, VGL, SPGL, ARGL による適応学年分布 n=2,136

学 年	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
					中	中	中	高	高	高	大	大	大	大			
					1	2	3	1	2	3	1	2	3	4			
度 数	3	8	6	4	4	18	37	79	166	303	491	505	327	146	33	3	3

Note: 学年は小学 1 年からの通し番号である。

4. 考察

TED 動画の適応学年 (GGL) は平均大学 1 年で, 大半が高校 2 から大学 4 年までだった。これは, 経験知とも重なっていて, 妥当であろう。教材を選ぶときの選定基準として GGL の採用を検討してもよいであろう。

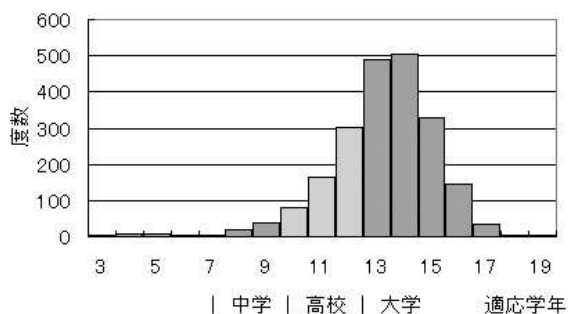


図 1. TED Talks の適応学年分布 n=2,136

5. 課題

課題を列举する。

- 1) 語彙, 発話速度, 構音速度の適応学年指標の精度向上と, 有効性の検証。
- 2) TED 動画の適応学年 (GGL) の活用法を考案。
- 3) 字幕ファイルから適応学年を返すウェブサイトの開発。

参考文献

mingle, (2016). <http://www.vector.co.jp/soft/winnt/edu/se506650.html>

田淵龍二・湯舟英一 (2016). 「音韻符号化の予測時間に基づく日本人英語学習者向けリーダビリティ公式の開発」. *Language Education & Technology*, 52, 359-388.