

## 観光英語検定の英文の計量言語学的解析 Metrical linguistic analysis of the Tourism English Proficiency Test

伴 浩美<sup>†</sup>  
Hiromi Ban

大藪 多可志<sup>‡</sup>  
Takashi Oyabu

### 1. はじめに

『平成 30 年版 観光白書』によると、2017 年 (平成 29 年)、1,789 万人の日本人が海外旅行に出掛け、2,869 万人の外国人が日本に観光に訪れている[1]。正に今、観光ブームと言えよう。それゆえ、観光に対する知識が益々重要とされ、また、世界共通語と言われる英語を使用する必要性が高まってきている。観光業の現場、すなわち、旅行・観光・ホテル・レストランサービス等の観光に関わる業務に必要とされる英語力の検定として、1989 年 (平成元年)に「観光英語検定」が始まった[2]。30 年目を迎え、これまでの受験者数は、延べ 20 万人以上に達している[2]。

本研究では、観光英検の英文にはどのような特徴がみられるのか、計量言語学的な解析を行った。すなわち、観光英検の英語について、文字種や単語種、及びその出現頻度を調査した。

### 2. 観光英語検定の概要

観光英検は、日本人旅行者を率いる旅行業者や添乗員、外国人旅行者に国内を案内するガイドなどを想定したコミュニケーション・接客能力の向上を目的としている[2]。

検定級は 1 級・2 級・3 級の各級に分かれ、それぞれ筆記 (Reading and Writing) およびリスニング試験が行われる。他の試験と同様 3 級から 1 級にかけて難易度が上がる。出題は、海外旅行および訪日外国人旅行における旅行・観光分野 (空港、交通、ホテル、ショッピング等) に絞り込まれ、一般的英語力に加えて、国内外の文化、地理、歴史の知識 (伝統芸能、郷土料理、自然景観等) を必要とする問題が含まれている[2]。

### 3. 解析試料と方法

本研究において解析した試料は、観光英語検定第 33 回 (2015 年 10 月実施) と第 35 回 (2015 年 10 月実施) の下記の試験問題である。

- 試料 1 : 第 33 回 第 1 級 Reading and Writing (T1Ra)
- 試料 2 : 第 35 回 第 1 級 Reading and Writing (T1Rb)
- 試料 3 : 第 33 回 第 2 級 Reading and Writing (T2Ra)
- 試料 4 : 第 35 回 第 2 級 Reading and Writing (T2Rb)
- 試料 5 : 第 33 回 第 3 級 Reading and Writing (T3Ra)
- 試料 6 : 第 35 回 第 3 級 Reading and Writing (T3Rb)
- 試料 7 : 第 33 回 第 1 級 Listening (T1La)
- 試料 8 : 第 35 回 第 1 級 Listening (T1Lb)
- 試料 9 : 第 33 回 第 2 級 Listening (T2La)
- 試料 10 : 第 35 回 第 2 級 Listening (T2Lb)
- 試料 11 : 第 33 回 第 3 級 Listening (T3La)
- 試料 12 : 第 35 回 第 3 級 Listening (T3Lb)

<sup>†</sup> 長岡技術科学大学 Nagaoka University of Technology

<sup>‡</sup> 日本海国際交流センター NIHONKAI International Exchange Center

なお、比較のため、日本の中学校教科書 *NEW HORIZON English Course 1, 2, 3* (2010, 東京書籍) (以下、「JHS 1, 2, 3」と記す) と高校教科書 *UNICORN ENGLISH COURSE I, II, READING* (2010, 文英堂) (以下、「HS 1, 2, 3」と記す) の本文の解析も行った。

解析プログラムは C++ で構成されている。このプログラムからは、各試料の文字と単語の頻度特性の他に、文の数、段落数、平均単語長など様々な情報が得られるよう配慮されている[3]。

### 4. 解析結果と考察

まず、各試料における使用頻度の高い文字の種類とその頻度を調べた。各試料の上位 50 位までを頻度の高い順に、縦軸は頻度の度合い、横軸は順位で、片対数でプロットし、その特性を

$$y = c * \exp(-bx) \quad (1)$$

で指数近似を行った[3]。得られた係数  $c$ 、 $b$  を Fig. 1 に示す。

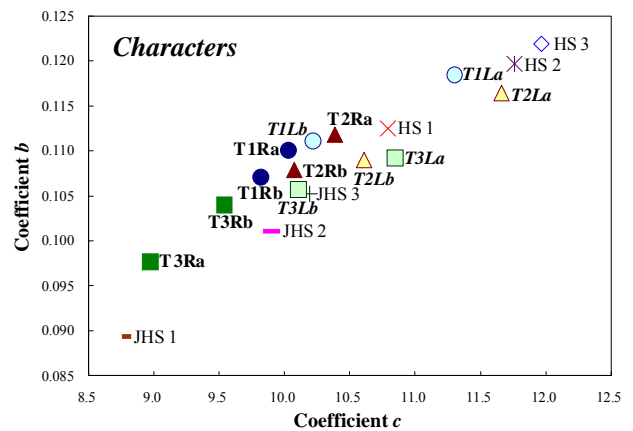


Fig. 1 Dispersions of coefficients  $c$  and  $b$  for character-appearance.

中高教科書を含む全ての試料の係数  $c$  と  $b$  にほぼニアな関係が見られ、教科書では学年が高くなるにつれて値が高くなっている。観光英検 3 級筆記の値が中学 1~2 年程度と低く、1・2 級リスニング第 33 回の試料が高校 2 年程度と高い。また、いずれの級もリスニングの方が筆記よりも高くなっている。前報において著者らは様々なジャンルの英文を解析し、上位 50 位までの文字について、近似式の係数  $c$  と  $b$  には正の相関が見られ、ジャーナリズムや技術英文に近いほど  $c$  と  $b$  の値が小さく、文学作品に近いほど、それらの値が大きい傾向にあることを示した[3]。従って、観光英検 3 級筆記や中学 1 年教科書はジャーナリズムや技術英文、また、観光英検リスニングや高校 3 年教科書は文学作品と同様の傾向があると言える。

次に単語頻度特性を調べた。観光英検について、tourist, world, park, room など観光関連の単語が出現頻度上位 20 位までに見られる。単語についても文字の場合と同様の解析を行ったところ、観光英検 2・3 級の筆記・リスニングの 8 試料の係数  $b$  の値が 0.0419~0.0466 の近い値となっており、内 6 試料は中学校教科書 3 試料 (0.0391~0.0425) よりも高く、高校教科書 3 試料 (0.0456~0.0479) よりも低い。なお、両係数について、観光英検 2・3 級筆記と高校 1・2 年教科書の全 4 試料が比較的近い値を取っており、一つのクラスターと見なすことが可能であると思われる。

次に、日本の中学校必修単語 508 語とアメリカの 4~8 歳児を対象とした *The American Heritage Picture Dictionary* (2003) に掲載されている 798 語 (以下、「基礎単語」と呼ぶ) を基に、単語の種類とその頻度から各試料の難易度を求めた。難易度を表すパラメータには、単語種数からの難易度 ( $D_{ws}$ ) と単語数からの難易度 ( $D_{wn}$ ) を考慮した。これらは、全単語数 ( $n_t$ )、全単語種数 ( $n_s$ )、必修[基礎]単語数 ( $n_{rs}$ )、各必修[基礎]単語数 ( $n(i)$ ) とすると、

$$D_{ws} = (1 - n_{rs} / n_s) \quad (2)$$

$$D_{wn} = \{ 1 - (1/n_t * \sum n(i)) \} \quad (3)$$

より求められる[3]。さらに適切な指数を与えるため、 $D_{ws}$  と  $D_{wn}$  を変数として主成分分析を行った[3]。分散共分散行列を用いて求めた第 1 主成分  $z$  は、必修、基礎単語ともに、 $[z = 0.7071 * D_{ws} + 0.7071 * D_{wn}]$  となった。これより得られた主成分得点をそれぞれ 1 次元で表したものを Fig. 2 に示す。図より、必修と基礎単語による値に正の相関が見られ、1 級は筆記・リスニングとも、高校 3 年教科書よりも難しく、2 級筆記は高校 1 年~2 年程度、他の 6 試料は中学 2 年~高校 1 年以下の難しさであることがわかる。また、いずれの級も筆記の方がリスニングより難易度が高いという特徴が見受けられる。

各試料のその他の計量的数値をまとめて Table 1 に記す。各試料における前置詞、関係詞等の使用頻度を求めたが、1 語ずつ意味を調べたわけではないため、前置詞、関係詞等とカウントしたものの中に、それ以外の品詞として用いられている単語も若干含まれている。まず、平均単語長は、1 級 4 試料が 5.754~6.041 語と、教科書 6 試料中最も長い高校 3 年の 5.568 語よりも長くなっている。次に、一文当たりの単語数は、いずれの級も筆記の方がリスニングよりも 0.205~4.732 語多い。1 級 4 試料は 19.938~23.307 語となっ

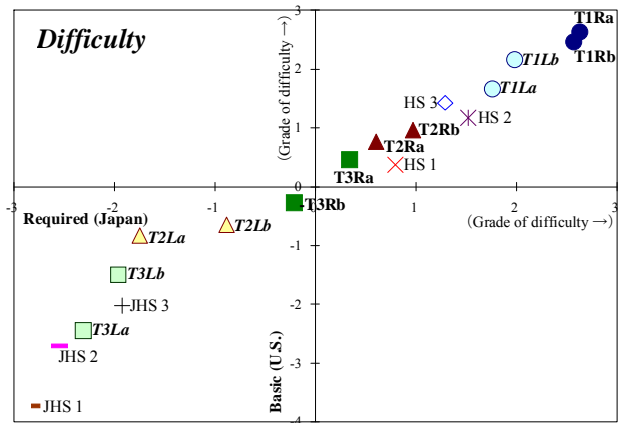


Fig. 2 Principal component scores of difficulty by required and basic vocabularies.

ており、高校 3 年の 15.778 語よりも多く、この点から考えても、1 級の英文は難しいものと思われる。また、助動詞の頻度については、1 級 4 試料が 0.612~0.876% であるのに対し、2・3 級 8 試料が 1.3161~2.444%、内 3 試料は 2% 以上となっており、1 級は断定的な表現が多く、2・3 級は、より多くの助動詞を用いて微妙なニュアンスを表していると言える。

### 5. おわりに

観光英語検定の問題の英文について、中学校・高校英語教科書と比較しながら、文字や単語の頻度特性を調べた。今後も観光英検の特徴抽出に関し、更に研究を重ねていくとともに、解析結果の教育への応用についても検討を行う予定である。

### 参考文献

- [1] 国土交通省 観光庁, “平成 30 年版 観光白書,” <http://www.mlit.go.jp/common/001260950.pdf> (2019/06/20 最終アクセス).
- [2] 全国語学ビジネス観光教育協会, “観光英語検定 - 英語の資格試験,” <http://kanko.zgb.gr.jp/index.html> (2019/06/20 最終アクセス).
- [3] Ban, H., H. Kimura and T. Oyabu, “Text Mining of English Materials for Business Management,” *International Journal of Engineering & Technical Research (IJETR)*, Vol. 3, Issue 8, pp. 238-243 (2015).

Table 1 Metrical data for each material.

	Tourism gr.1, R (a)	Tourism gr.1, R (b)	Tourism gr.2, R (a)	Tourism gr.2, R (b)	Tourism gr.3, R (a)	Tourism gr.3, R (b)	Tourism gr.1, L (a)	Tourism gr.1, L (b)	Tourism gr.2, L (a)	Tourism gr.2, L (b)	Tourism gr.3, L (a)	Tourism gr.3, L (b)	JHS 1 (JHSzen 1)	JHS 2 (JHSzen 2)	JHS 3 (JHSzen 3)	HS 1 (HSzen 1)	HS 2 (HSzen 2)	HS 3 (HSzen 3)
Total num. of characters	26,609	26,205	13,527	13,950	10,667	10,779	22,935	20,545	10,257	9,323	7,442	7,529	6,824	14,362	13,387	44,279	67,662	88,289
Total num. of character-type	74	75	75	76	74	72	74	73	67	67	64	69	69	69	71	73	75	76
Total num. of words	4,405	4,512	2,395	2,507	1,931	1,933	3,986	3,509	1,923	1,726	1,444	1,428	1,339	2,876	2,594	8,083	12,264	15,857
Total num. of word-type	1,684	1,683	962	1,011	747	750	1,550	1,343	695	671	520	568	497	799	764	2,059	2,657	3,594
Total num. of sentences	189	224	158	171	146	141	195	176	184	152	170	146	251	394	317	633	890	1,005
Total num. of paragraphs	60	59	71	81	81	77	62	43	88	89	97	82	233	227	177	163	261	260
Mean word length	6.041	5.808	5.648	5.564	5.524	5.576	5.754	5.855	5.334	5.402	5.154	5.272	5.096	4.994	5.161	5.478	5.517	5.568
Words/sentence	23.307	20.143	15.158	14.661	13.226	13.709	20.441	19.938	10.451	11.355	8.494	9.781	5.335	7.299	8.183	12.769	13.780	15.778
Sentences/paragraph	3.150	3.797	2.225	2.111	1.802	1.831	3.145	4.093	2.091	1.708	1.753	1.780	1.077	1.736	1.791	3.883	3.410	3.865
Commas/sentence	1.180	1.170	0.608	0.626	0.671	0.617	1.144	1.000	0.342	0.513	0.300	0.315	0.263	0.223	0.331	0.694	0.801	0.977
Repetition of a word	2.616	2.681	2.490	2.480	2.585	2.577	2.572	2.613	2.767	2.572	2.777	2.514	2.694	3.599	3.395	3.926	4.616	4.412
Freq. of prepositions (%)	16.030	15.402	13.366	14.723	12.949	12.257	14.854	15.950	13.676	14.196	12.259	12.603	9.110	11.788	12.188	14.769	14.810	15.052
Freq. of relatives (%)	1.611	1.551	1.671	1.517	1.450	2.018	1.854	1.877	1.716	1.507	1.455	2.170	1.792	1.392	1.927	1.745	2.421	2.383
Freq. of auxiliaries (%)	0.612	0.842	1.463	1.316	2.072	1.605	0.876	0.654	2.444	1.971	1.593	2.240	0.897	1.530	1.119	0.802	1.215	1.217
Freq. of pers. pronouns (%)	2.430	2.637	7.898	5.984	6.838	6.935	4.690	3.045	11.284	9.270	10.045	9.172	17.476	15.511	10.684	9.324	8.707	8.393