

英語検定の英文の計量的文体解析
Metrical linguistic analysis of English proficiency tests

伴 浩美[†]
Hiromi Ban

大藪 多可志[‡]
Takashi Oyabu

1. はじめに

従来の大学入試センター試験に代わり、2021年1月に実施される大学入学共通テストでは、英語4技能「読む・聞く・話す・書く」を測定するため、英語民間試験が導入される予定だった。しかし、2019年11月1日、4技能評価を支援する「大学入試英語成績提供システム」の導入を見送ることが発表された[1]。

予定通り実施するかどうかに関しては、高校生をはじめ多数の人から、賛成・反対、さまざまな意見があったところであるが、英語教育充実のために導入を予定してきた試験を経済的な状況や居住している地域に関わらず、等しく安心して受けられるようにするためには、更なる時間が必要だと判断、今回は導入が見送られた[1]。

本研究では、導入される予定だった民間試験の幾つかについて、それら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、英語民間試験問題の英語について、文字種や単語種、及びその出現頻度を調査した。

2. 解析試料と方法

本研究において解析した試料を Table 1 に示す。

Table 1 Metrical data for each material.

No.	試験名称	略	試料
1.	英検 2級	EI	Practical English Proficiency Grade 2, 2017-2
2.	TEAP	TE	TEAP実践問題集(旺文社, 2015)の「TEAP 見本問題 1」
3.	IELTS	IE	"Test 1" in <i>IELTS 14 Academic Student's Book with Answers with Audio: Authentic Practice Tests (IELTS Practice Tests)</i> (Cambridge University Press, 2019)
4.	TOEIC L&R	IC	"Test 1" in <i>Official TOEIC Listening & Reading Tests 3</i> (Educational Testing Service, 2017)
5.	TOEFL iBT	FL	"Authentic TOEFL iBT Practice Test 1" in <i>The Official Guide to the TOEFL Test, Fourth Edition</i> (Educational Testing Service, 2012)
6.	GTEC CBT	GT	GTEC 公式問題集 CBT 第2版(ベネッセコーポレーション, 2018)の「本試験1回目」
7.	ケンブリッジ英検	CA	ケンブリッジ英検 PET 実践問題集(Z会, 2016)の「模擬問題」

いずれも筆記試験問題について解析を行った[2]。

なお、比較のため、大学入試センター試験(本試験)

[†] 長岡技術科学大学 Nagaoka University of Technology

[‡] 日本海国際交流センター NIHONKAI International Exchange Center

「英語(筆記)」2016年1月実施~2020年1月実施の5回分(以下、「C16, C17, C18, C19, C20」と記す)、また、日本の中学校教科書 *NEW HORIZON English Course 1, 2, 3* (東京書籍, 2010)(以下、「JHS 1, 2, 3」と記す)と高校教科書 *UNICORN ENGLISH COURSE I, II, READING* (文英堂, 2010)(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]。

3. 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。

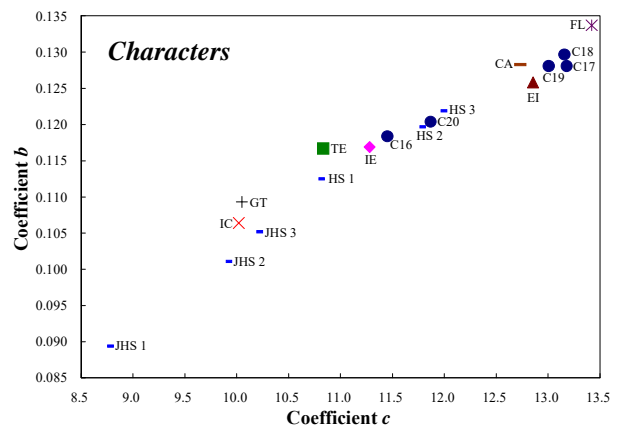


Fig. 1 Dispersions of coefficients c and b for character-appearance.

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る。教科書では学年が高くなるにつれて値が高くなっており、TOEICやGTECの値は中学3年程度と低い。一方TOEFLが最も高く、センター試験の近年3年の試料がそれに続いている。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、TOEICやGTECはジャーナリズムや技術英文、一方、TOEFLは文学作品と同様の傾向があると言える。

続いて単語頻度特性を調べた。TEAP と GTEC では students, IELTS では employees, work といった単語が出現頻度上位 30 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、係数 c については、全 18 試料中 TEAP(1.7669) が最も低く、GTEC (2.1703) が最も高い。TOEFL (1.9707) とセンター試験の内 3 試料 (1.9236~1.9465) が平均値 1.9229 に近づいている。なお、両係数について、英検 2 級とセンター試験の内 2 試料が比較的近い値を取っている。

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]。更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、基礎単語共に、 $[z = 0.7071 * D_{ws} + 0.7071 * D_{wn}]$ 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、必修単語と基礎単語による値に正の相関が見られ、TOEFL, IELTS, TOEIC, TEAP は高校 3 年教科書よりも難しく、一方ケンブリッジ英検は中学 3 年教科書に次いで易しいことがわかる。また、センター試験 5 試料全てが基礎単語による高校 1 年~2 年の間の近い値を取っており、それらより若干低い値の英検 2 級と共に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。

各試料のその他の計量的数値をまとめて Table 2 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、平均単語長について、民間試験ではケンブリッジ英検が高校 1 年教科書より短い一方 TOEFL, IELTS, TEAP が 6 語を超えており、全 7 試料平均 5.840、標準偏差 0.317 と、センター試験 5 試料平均 5.644、標準偏差 0.129 よりも大きな値となっている。

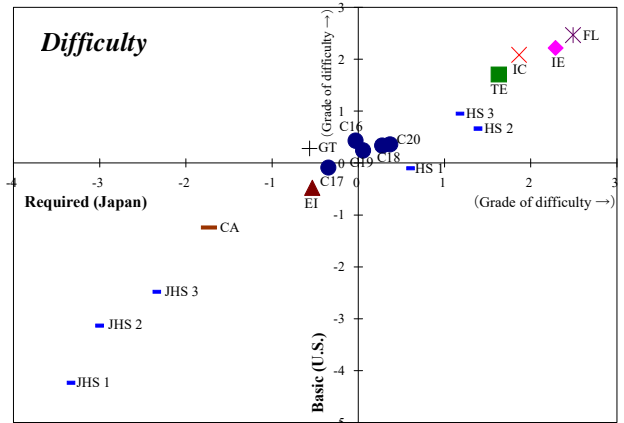


Fig. 2 Principal component scores of difficulty by required and basic vocabularies.

一文当たりの単語数についても、民間試験の平均 18.459、標準偏差 4.881、センター試験の平均 14.925、標準偏差 0.654 と、民間試験の試料の違いによる差が大きいことが窺われる。前置詞、関係詞等の頻度についても同様であり、標準偏差が民間試験は 0.582~2.353、センター試験は 0.331~0.998 で、民間試験の方が 0.251 (前置詞) ~1.355 (人称代名詞) 高く、バラツキが大きいことが明らかとなった。

4. おわりに

英語民間試験問題の英文について、センター試験問題や中学校・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。今後も英語試験問題の特徴抽出に関し、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] 株式会社イード, “【大学受験】英語民間試験の見送り決定、2024 年度より導入,” Resemom, <https://resemom.jp/article/2019/11/01/53171.html> (2020/06/18 最終アクセス).
- [2] 旺文社, “『英語外部検定利用入試』特集,” 大学受験パスナビ, <https://passnavi.evidus.com/gaibukentei/index.html> (2020/06/18 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 2 Metrical data for each material.

	EIKEN gr. 2	TEAP	IELTS	TOEIC	TOEFL	GTEC	Cam.	Center 2016	Center 2017	Center 2018	Center 2019	Center 2020	JHS 1 (Horizon 1)	JHS 2 (Horizon 2)	JHS 3 (Horizon 3)	HS 1 (Uncora 1)	HS 2 (Uncora 2)	HS 3 (Uncora 3)
Total num. of characters	11,157	19,489	17,861	21,234	12,215	25,326	8,870	17,808	17,825	18,405	17,481	17,554	6,824	14,362	13,387	44,279	67,662	88,289
Total num. of character-type	69	73	74	78	68	77	63	71	71	74	71	74	69	69	71	73	75	76
Total num. of words	1,984	3,241	2,878	3,654	1,945	4,523	1,658	3,100	3,199	3,172	3,088	3,228	1,339	2,876	2,594	8,083	12,264	15,857
Total num. of word-type	737	1,226	1,079	1,375	806	1,117	650	1,067	1,043	1,107	1,067	1,116	497	799	764	2,059	2,657	3,594
Total num. of sentences	131	180	98	247	95	264	116	203	220	198	211	228	251	394	317	633	890	1,005
Total num. of paragraphs	42	64	39	123	18	109	36	73	78	69	67	75	233	227	177	163	261	260
Mean word length	5.623	6.013	6.206	5.811	6.260	5.599	5.350	5.745	5.572	5.802	5.661	5.438	5.096	4.994	5.161	5.478	5.517	5.568
Words/sentence	15.145	18.006	29.367	14.794	20.474	17.133	14.293	15.271	14.541	16.020	14.635	14.158	5.335	7.299	8.183	12.769	13.780	15.778
Sentences/paragraph	3.119	2.813	2.513	2.008	5.278	2.422	3.222	2.781	2.821	2.870	3.149	3.040	1.077	1.736	1.791	3.883	3.410	3.865
Commas/sentence	0.809	0.933	1.653	0.518	1.211	0.708	0.405	0.768	0.686	0.732	0.820	0.728	0.263	0.223	0.331	0.694	0.801	0.977
Repetition of a word	2.692	2.644	2.667	2.657	2.413	4.049	2.551	2.905	3.067	2.865	2.894	2.892	2.694	3.599	3.395	3.926	4.616	4.412
Freq. of prepositions (%)	14.665	14.782	15.116	14.201	15.576	13.728	15.197	14.711	15.071	14.257	15.059	14.409	9.110	11.788	12.188	14.769	14.810	15.052
Freq. of relatives (%)	2.872	2.715	3.336	1.640	3.186	1.967	2.109	2.676	2.750	3.090	2.267	2.169	1.792	1.392	1.927	1.745	2.421	2.383
Freq. of auxiliaries (%)	1.410	1.110	0.904	2.436	1.542	3.293	1.930	1.293	2.345	1.797	1.523	1.519	0.897	1.530	1.119	0.802	1.215	1.217
Freq. of personal pronouns (%)	6.801	4.567	4.932	6.840	2.311	8.950	9.528	7.228	10.068	8.800	8.744	9.791	17.476	15.511	10.684	9.324	8.707	8.393