

大学入学共通テスト英語の計量言語学的解析 Metrical linguistic analysis of English for the Common Test for University Admissions

伴 浩美[†]
Hiromi Ban

大藪 多可志[‡]
Takashi Oyabu

1. はじめに

独立行政法人「大学入試センター」と各大学との共同により、およそ30年にわたり実施されてきた「大学入試センター試験」(National Center Test for University Admissions) (以下、「センター試験」と記す)の後継として「大学入学共通テスト」(Common Test for University Admissions) (以下、「共通テスト」と記す)が2021年度入試から実施された。解答方法はセンター試験同様マーク式のみであるが、思考力・判断力・表現力等を一層重視した評価ができるよう、作問や出題形式が見直された[1]。中でも英語については出題傾向の変化が話題になっている。「聞く・話す・読む・書く」の4技能を測る民間試験の活用が見送られた一方で、実用英語やリスニングの重視が鮮明になった[2]。

本研究では、共通テスト英語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、共通テスト英語リーディングとリスニングの英語について、文字種や単語種、及びその出現頻度を調査した。

2. センター試験と共通テスト英語の比較

共通テスト英語について2020年度センター試験との相違点は下記の通りである[2]。

《配点》 筆記 200点・リスニング 50点→リーディング 100点・リスニング 100点

《リーディング》

- ・試験時間 80分→変更なし
- ・設問数 大問6問→変更なし
- ・設問 日本語→英語
- ・発音, アクセントの単独問題, 語句整序の出題あり→なし
- ・出題の特徴 発音から読解まで総合的に→すべて読解問題. SNS, プレゼンテーションなど実用的な言語活動を意識した場面設定や複数情報の整理・比較

《リスニング》

- ・試験時間 60分(うち30分が解答時間)→変更なし
- ・設問数 大問4問→6問
- ・音声は2回流れる→6問中4問が「1回読み」
- ・出題の特徴 音声の内容理解が中心→グラフやワークシートを用い, 米・英ほか多様な話者に

3. 解析試料と方法

本研究において解析を行った試料は、2021年度共通テスト第1日程・第2日程英語(リーディング, リスニング) (以下、「C21R1, C21R2, C21L1, C21L2」と記す)計4試料である。なお、比較のため、センター試験2016

～2020年度英語(筆記, リスニング) (以下、「N16R, N17R, ... N19L, N20L」と記す)10試料, 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]。

4. 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。

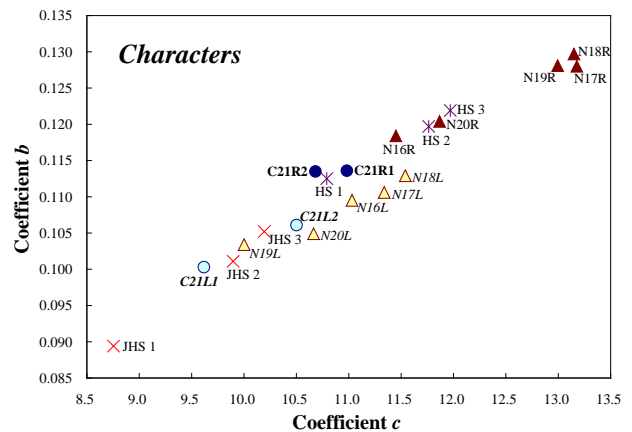


Fig. 1 Dispersions of coefficients c and b for character-appearance.

中高教科書を含む全ての試料の係数 c と b にほぼリ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いる。C21R1, C21R2 は HS 1 と同程度, C21L1, C21L2 は HS 1 より値が低い。一方、センター試験筆記全5試料は HS 1 よりも高い値となっている。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、共通テストの英文はジャーナリズムや技術英文と同様の傾向があると言える。

[†] 三条市立大学 Sanjo City University

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

単語についても同様の解析を行ったところ、全 20 試料の係数 c と b に弱い正の相関が見られ、C21L1 が最も低い値となっている。係数 c について、共通テスト 4 試料の値が N16R を除く他の 15 試料よりも低い。係数 b については、リスニング全 7 試料が 0.0347~0.0399、リーディング全 7 試料が 0.0402~0.0447、高校教科書が 0.0456~0.0479 と順に高くなっていっている。なお、両係数について、C21R1、C21L2、N16R、N19R、JHS 3 の 5 試料は比較的近い値を取っており、1 つのクラスタ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]。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難易度には正の相関が見られる。C21L2 が共通テスト 4 試料中最も低くなっているものの、センター試験リスニング 4 試料よりも難しい。C21L1 は N18R と同程度で C21R1 より難しく、C21R2 は試験 14 試料中最も難しく、HS 2・HS 3 と同程度の難易度であることが明らかとなった。

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、平均単語長について、C21R2 が 5.734 字で、全 20 試料中 N18R、N16R に次いで 3 番目に長い、C21R1 と共通テストリスニング 2 試料は 7~9 番目に長く、これらは HS 2 より長く、HS 3 よりも短く、かつ、センター試験リスニング全 5 試料と N18R よりも長いものとなっている。一文当たりの

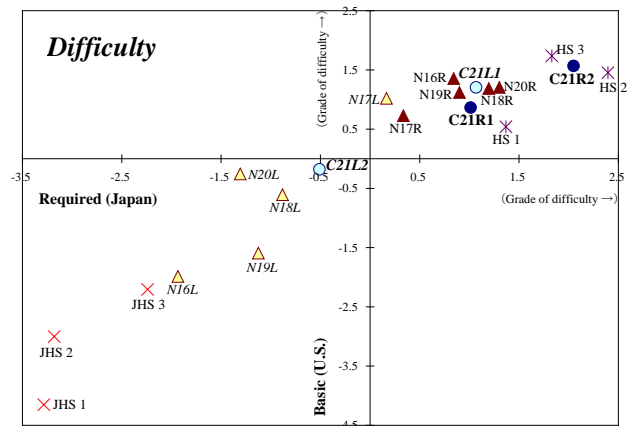


Fig. 2 Principal component scores of difficulty by required and basic vocabularies.

単語数について C21R2 は 15.414 語であり、N18R の 16.020 語、HS 3 の 15.778 語に次いで 3 番目に多く、この点からも C21R2 の英文が難しいことが窺われる。なお、助動詞の頻度については C21L2 が 3 番目に高く、助動詞を多く用いて微妙なニュアンスを表していることが窺われる。さらに人称代名詞の頻度については、C21R2 が全 20 試料中最も少なく、C21R1 も N16R に次いで 3 番目に少ないという特徴が見受けられた。

5. おわりに

2021 年度共通テスト英語について、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、さら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] 河合塾 Kei-Net, “大学入試の仕組みを理解しよう！共通テストについて理解しよう,” <https://www.keinet.ne.jp/exam/basic/structure/center.html> (2021/06/13 最終アクセス)。
- [2] 朝日新聞デジタル, “「使える英語」 出題内容が様変わり 共通テスト、懸念も,” <https://www.asahi.com/articles/ASP1T6SKHP1LUTIL05Y.html> (2021/06/13 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.

	Com. R1	Com. R2	NatL. 16, R	NatL. 17, R	NatL. 18, R	NatL. 19, R	NatL. 20, R	Com. LI	Com. L2	NatL. 16, L	NatL. 17, L	NatL. 18, L	NatL. 19, L	NatL. 20, L	JHS 1	JHS 2	JHS 3	HS 1	HS 2	HS 3
Total num. of characters	20,011	17,501	17,808	17,825	18,405	17,481	17,554	8,478	8,455	6,071	6,453	6,397	6,220	6,247	6,824	14,362	13,387	44,279	67,662	88,289
Total num. of character-type	76	76	71	71	74	71	74	67	66	67	68	64	64	67	69	69	71	73	75	76
Total num. of words	3,594	3,052	3,100	3,199	3,172	3,088	3,228	1,523	1,519	1,169	1,175	1,179	1,217	1,182	1,339	2,876	2,594	8,083	12,264	15,857
Total num. of word-type	1,178	1,106	1,067	1,043	1,107	1,067	1,116	679	644	496	530	549	503	508	497	799	764	2,059	2,657	3,594
Total num. of sentences	237	198	203	220	198	211	228	141	137	100	107	113	136	123	251	394	317	633	890	1,005
Total num. of paragraphs	73	54	73	78	69	67	75	96	99	94	92	89	97	96	233	227	177	163	261	260
Mean word length	5.568	5.734	5.745	5.572	5.802	5.661	5.438	5.567	5.566	5.193	5.492	5.426	5.111	5.285	5.096	4.994	5.161	5.478	5.517	5.568
Words/sentence	15.165	15.414	15.271	14.541	16.020	14.635	14.158	10.801	11.088	11.690	10.981	10.434	8.949	9.610	5.335	7.299	8.183	12.769	13.780	15.778
Sentences/paragraph	3.247	3.667	2.781	2.821	2.870	3.149	3.040	1.469	1.384	1.064	1.163	1.270	1.402	1.281	1.077	1.736	1.791	3.883	3.410	3.865
Commas/sentence	0.992	0.803	0.768	0.686	0.732	0.820	0.728	0.617	0.701	0.720	0.776	0.664	0.493	0.593	0.263	0.223	0.331	0.694	0.801	0.977
Repetition of a word	3.051	2.759	2.905	3.067	2.865	2.894	2.892	2.243	2.359	2.357	2.217	2.148	2.419	2.327	2.694	3.599	3.395	3.926	4.616	4.412
Freq. of prepositions (%)	14.359	14.716	14.711	15.071	14.257	15.059	14.409	11.559	13.760	11.467	13.957	12.471	11.502	11.255	9.110	11.788	12.188	14.769	14.810	15.052
Freq. of relatives (%)	1.835	1.770	2.676	2.750	3.090	2.267	2.169	2.627	2.567	3.423	4.169	2.799	2.793	2.961	1.792	1.392	1.927	1.745	2.421	2.383
Freq. of auxiliaries (%)	1.946	1.147	1.295	2.345	1.797	1.523	1.519	1.773	2.172	2.224	2.042	1.782	1.890	2.115	0.897	1.530	1.119	0.802	1.215	1.217
Freq. of personal pronouns (%)	8.237	7.048	7.228	10.068	8.800	8.744	9.791	10.048	10.070	14.804	12.340	12.642	14.295	14.147	17.476	15.511	10.684	9.324	8.707	8.393