

統計処理による学習対象の最適化

熊本県立湧心館高等学校 西村 信一

1 はじめに

近年、ビッグデータ・AIなどの活用が理工学的な分野に留まらず多くの社会現象への応用も顕著である。昨年、高校数学部会長の月井先生のご尽力により、熊本大学との連携による統計解析分野の研究を主体とする本事業を推進することとなった。今回、身近のテーマである入試問題を対象とする報告を行う。特に過去問を分析することにより最適な学習教材(学習対象)を提供するとともに、ICT教材としても活用できるものを目指す。

平成2年(1990年)に導入されたセンター試験も令和2年(2020年)を最後に時代の変遷とともにその役目を終える。これまでの30年間(本試・追試)で実施された60回の試験問題は膨大なものとなった。来年は最後の試験であると同時に過去の出題傾向を最も活用できる試験でもある。統計的な手法を用いて最適な戦略(strategy)の構築を目指すものである。

2 検証データについて

全ての教科について、出題の特徴や傾向を分析することは困難であるため、毎回、出題されている問題に着目し、60回すべての試験で出題された英語のアクセントの問題¹について検証する。

2019 本試

第一アクセント(第一強勢)の位置がほかの三つと異なるものを、それぞれ下の①~④のうちから一つずつ選べ。

① agree ② control ③ equal ④ refer

この出題形式(4問×4択)は2013年以降続いている。問題から分かるように、4つすべての正しいアクセントの位置が分からなくとも、3つ分かれば正解を見つけ出すことができる。仮にどれか2つだ

け正しいアクセント位置を知っていれば、その確率(条件つき確率)は

$$\frac{1}{2}$$

となる。そこで正解率を上げるために、最も効率的な戦略について考察を行う。

アクセント問題として出題された全548語の出題回数は次のとおりである。

回数	1	2	3	4	計
語数	413	109	22	4	548

- 4回 [commúnitý, énergý, entertáin, succéss](#)
 3回 [ác-cu-ra-cý, ad-míre, ár-chi-tec-ture, ca-réer, cón-cen-trate, con-tín-ue, de-móc-ra-cý, dém-on-strate, ex-pér-i-ment, ín-di-cate, ín-dus-try, mys-té-ri-ous, of-fí-cial, par-tíc-i-pate, per-cént, phi-lós-o-phy, póli-tics, pre-fér, sát-is-fý, sci-en-tíf-ic, tech-nól-o-gý, tún-nel](#)

このように再度出題される単語も少なくない。近年の本試験(2017-2019)を出題回数および出題法則(次ページを参照)の観点から検証する。

2019年(本試験)	出題歴	法則
(1) ① a-grée	初回	22
② con-tról	91本	
③ é-qual	初回	13
④ re-fér	初回	
(2) ① ap-próv-al	初回	13
② cál-en-dar	09本	
③ re-mém-ber	93本	
④ suc-céss-ful	初回	4
(3) ① chár-ac-ter	99追	
② de-lí-cious	08本	11
③ óp-po-site	07本	10
④ trág-e-dý	初回	

¹<http://kumamoto.s12.xrea.com/lis.html>

(4)	①	ár-chi-tec-ture	09 追,13 本	
	②	bi-ól-o-gy	初回	19
	③	spec-tác-u-lar	初回	14
	④	sur-prís-ing-ly	初回	4

2018 年 (本試験)		出題歴	法則	
(1)	①	ad-ván-ce	初回	33
	②	dán-ger	初回	
	③	én-gine	初回	25
	④	lím-it	初回	
(2)	①	de-pós-it	初回	5
	②	foun-dá-tion	初回	11
	③	o-pín-ion	90 本	
	④	rég-is-ter	初回	
(3)	①	á-gen-cy	初回	
	②	fré-quent-ly	初回	4
	③	in-tro-dú-ce	初回	
	④	óf-fi-cer	初回	4
(4)	①	chám-pi-on-ship	初回	2
	②	de-lív-er-y	初回	5
	③	rél-a-tive-ly	初回	4
	④	sú-per-mar-ket	92 追	2

2017 年 (本試験)		出題歴	法則	
(1)	①	ma-rí-ne	初回	25
	②	ráp-id	初回	
	③	se-vére	初回	
	④	u-ní-que	初回	27
(2)	①	e-nór-mous	初回	18
	②	év-i-dence	初回	32
	③	sát-el-lite	初回	10
	④	týp-i-cal	初回	12
(3)	①	as-sém-bly	初回	4
	②	cor-re-spónd	初回	
	③	dis-tín-guish	初回	5
	④	ex-pén-sive	90 追	12
(4)	①	déf-i-nite-ly	初回	4
	②	dem-o-crát-ic	90 追	12
	③	in-de-pénd-ence	初回	32
	④	res-o-lú-tion	初回	11

過去 30 年 (1990–2019 本試・追試) における主な基本法則に該当する出題回数および出題語数は次のとおりである。

基本法則		回	語
1	1 音節	0	0
2	2 語の合成語	23	19
3	名前動後	14	11
4	派生語	34	28
5	be-, de-, dis-, pre-, un-	21	16
6	by-, pan-, para-	0	0
7	no-, any-, some-, every-	0	0
8	-ever, -self, -ese, -out	1	1
9	-ate	43	30
10	-ite, -ude, -ute, -tribute	9	6
11	シャミセン接尾辞	41	33
12	-ic, -ics, -ical, -sive	47	38
13	-ial, -ual, -al	28	21
14	-iar, -ior, -ilar, -ular	7	4
15	-icle, -ium, -ient, -ience, -ean	4	3
16	-ity, -ety, -ty	18	12
17	-ify, -efy, -fy	4	2
18	-my, -phy, -thy, -sis, -ous	19	13
19	-cracy, -logy, -meter	10	5
20	-gram, -graph, -grapher	5	3
21	-ism	0	0
22	-ee, -een, -eer, -oo, -oon	13	8
23	-ay, -oy	3	2
24	-ade, -end, -pose	5	4
25	-ine, -ice	10	9
26	-ise, -ize	8	7
27	-esque, -ique, -igue, -ette	5	4
28	-able, -ible	7	7
29	-age	7	6
30	-ain	6	3
31	-tive	10	10
32	-ent, -ence	56	46
33	-ant, -ance	26	21
合計		484	372
出題総数		702	548
合計/出題総数 (%)		68.9	67.8

3 正解数の期待値

過去7年(2013-2019)の本試験における出題状況について調査を行った。過去に出題された(出題年以前)単語の集合を A , 基本法則に分類できる単語の集合を B とすると, 次の結果が判明した。

2019年	(1)	(2)	(3)	(4)
$n(A)$	1	2	3	1
$n(B)$	2	2	2	3
$n(A \cup B)$	3	4	3	4

2018年	(1)	(2)	(3)	(4)
$n(A)$	0	1	0	1
$n(B)$	2	2	2	4
$n(A \cup B)$	2	3	2	4

2017年	(1)	(2)	(3)	(4)
$n(A)$	0	0	1	1
$n(B)$	2	4	3	4
$n(A \cup B)$	2	4	3	4

2016年	(1)	(2)	(3)	(4)
$n(A)$	2	1	2	2
$n(B)$	1	3	3	3
$n(A \cup B)$	2	3	3	3

2015年	(1)	(2)	(3)	(4)
$n(A)$	2	1	0	2
$n(B)$	0	3	3	3
$n(A \cup B)$	2	3	3	3

2014年	(1)	(2)	(3)	(4)
$n(A)$	3	1	2	0
$n(B)$	2	3	3	3
$n(A \cup B)$	4	3	4	3

2013年	(1)	(2)	(3)	(4)
$n(A)$	4	2	2	3
$n(B)$	2	3	3	3
$n(A \cup B)$	4	4	4	4

出題された4語の中で他の3語と第一アクセントの位置が異なるものを答える問題であるから, 4語の中でアクセントの位置を知っている語数を i とすると, 正解する確率 $P_i (i = 0, 1, 2, 3, 4)$ は

$$P_0 = P_1 = \frac{1}{4}, \quad P_2 = \frac{1}{2}, \quad P_3 = P_4 = 1$$

センター試験で過去に出題された単語のアクセントの位置をすべて知っているとき, 全問4題中の正解数の期待値 E_A は, 次のようになる。この7年の E_A の平均は1.75であるため, 過去問中心の学習の成果は乏しい(正解数の期待値 ≥ 1)。

$$E_A(2019年) = \frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{4} = 2$$

$$E_A(2018年) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$$

$$E_A(2017年) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$$

$$E_A(2016年) = \frac{1}{2} + \frac{1}{4} + \frac{1}{2} + \frac{1}{2} = 1.75$$

$$E_A(2015年) = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{2} = 1.5$$

$$E_A(2014年) = 1 + \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 2$$

$$E_A(2013年) = 1 + \frac{1}{2} + \frac{1}{2} + 1 = 3$$

前ページの基本33法則についてのみ学習した場合, 正解数の期待値 E_B は, 次のようになる。7年間の E_B の平均値は $\frac{22}{7} = 3.14\dots$ 。

$$E_B(2019年) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 1 = 2.5$$

$$E_B(2018年) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 1 = 2.5$$

$$E_B(2017年) = \frac{1}{2} + 1 + 1 + 1 = 3.5$$

$$E_B(2016年) = \frac{1}{4} + 1 + 1 + 1 = 3.25$$

$$E_B(2015年) = \frac{1}{4} + 1 + 1 + 1 = 3.25$$

$$E_B(2014年) = \frac{1}{2} + 1 + 1 + 1 = 3.5$$

$$E_B(2013年) = \frac{1}{2} + 1 + 1 + 1 = 3.5$$

基本33法則を学習した上で, 過去問も学習していたとすると, 正解数の期待値 $E_{A \cup B}$ は大きく, 7年間の $E_{A \cup B}$ の平均値は $\frac{51}{14} = 3.64\dots$ となる。

$$E_{A \cup B}(2019年) = 4, \quad E_{A \cup B}(2018年) = 3,$$

$$E_{A \cup B}(2017年) = 3.5, \quad E_{A \cup B}(2016年) = 3.5,$$

$$E_{A \cup B}(2015年) = 3.5, \quad E_{A \cup B}(2014年) = 4,$$

$$E_{A \cup B}(2013年) = 4$$

30年間60回(本試・追試)の試験で出題された単語の種類は548語あり, そのうち基本33法則に該当する単語は372語で全体の67.9%に相当する。過去問よりも基本法則を中心に学習した方が効率的であることが判明した。

過去 30 年 60 回の試験から例外が少なく、出題率の高い法則を 3 つ (9,11,12) だけ紹介する。

9. -ate で終わる 3 音節以上の単語は、その 2 つ前の音節にアクセントがある。基本的に例外はなく、出題回数は 43 回。なお、動詞は [eit], 形容詞は [ot/it] と発音される。

ín-di-cate (2007 本試, 2017 追試)
ád-e-quate (2008 本試, 2013 追試)

11. -cian, -sian, -(s)sion, -(s)sion, -gion, -(s)cious, -tious, -xious, -gious, -geous, -cial, -sial, -tial, -cient, -tient, -tience, -science で終わるものをシャミセン接尾辞と総称し、これらの接尾辞の直前の音節にアクセントがあり、ほとんど例外もない。出題回数は 41 回。

de-cí-sion (2016 本試)
com-pe-tí-tion (1994 追試, 2013 追試)
re-lí-gion (1991 追試, 2012 本試)
de-lí-cious (2008 本試, 2019 本試)
am-bí-tious (2015 本試)
fi-nán-cial (2010 追試, 2019 追試)
suf-fí-cient (2015 本試)

12. -ic, -ics, -ical, -sive で終わる単語は、その直前の音節にアクセントがある。出題回数は 47 回。例外で出題されたのは、**pól-i-tics**(1996 追試, 2007 追試, 2016 本試) のみ。

dy-nám-ic (2013 本試)
math-e-mát-ics (2008 本試)
gram-mát-i-cal (1993 追試, 2007 本試)
ex-pén-sive (1990 追試, 2017 本試)

これら 3 つの法則に該当する単語で直近 7 年の本試験で出題されたものが次である。

2019 年 **de-lí-cious**
 2018 年 **foun-dá-tion**
 2017 年 **týp-i-cal**, **dem-o-crát-ic**, **res-o-lú-tion**,
ex-pén-sive
 2016 年 **de-cí-sion**, **pól-i-tics**, **po-tén-tial**
 2015 年 **am-bí-tious**, **suf-fí-cient**, **ap-pró-pri-ate**
 2014 年 **e-lím-i-nate**, **in-vés-ti-gate**, **do-més-tic**
 2013 年 **es-sén-tial**, **dy-nám-ic**, **his-tór-i-cal**

4 データ分析に基づく学習戦略

アクセント問題ばかりを勉強することは現実的でない。例えば、過去に 10 回以上出題された 16 法則に限って学習したとき、これらの法則に該当する単語の出題回数の総数は 413 回、単語の出題総数が 702 回であるから、ある単語がこれらの法則に該当する確率 p は

$$p = \frac{413}{702}$$

1 題 4 語の単語のうち、16 法則に該当する単語の個数を X とすると ($q = 1 - p$)

X	0	1	2	3	4
$P(X)$	q^4	$4pq^3$	$6p^2q^2$	$4p^3q$	p^4
正解率	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1	1

正解率の期待値を E とすると

$$\begin{aligned} E &= \frac{1}{4} \cdot q^4 + \frac{1}{4} \cdot 4pq^3 + \frac{1}{2} \cdot 6p^2q^2 + 1 \cdot 4p^3q + 1 \cdot p^4 \\ &= \frac{1}{4}(q^4 + 4pq^3 + 6p^2q^2 + 4p^3q + p^4) \\ &\quad + \frac{3}{4}p^2(2q^2 + 4pq + p^2) \\ &= \frac{1}{4}(p+q)^4 + \frac{3}{4}p^2\{2(p+q)^2 - p^2\} \\ &= \frac{1}{4} + \frac{3}{4}p^2(2-p^2) = 1 - \frac{3}{4}(1-p^2)^2 \end{aligned}$$

したがって

$$p = \frac{413}{702} \text{ より } E \doteq 0.679$$

7 回以上出題された 21 法則まで学習すると

$$p = \frac{451}{702} \text{ より } E \doteq 0.741$$

すべての基本法則を学習しても 8 割程度に留まり、高得点になるにつれ、学習の効率は悪くなる。

$$p = \frac{484}{702} \text{ より } E \doteq 0.794$$

5 データ分析に基づく ICT 教材

本分析による教材「アクセント基本法則」

<http://kumamoto.s12.xrea.com/accent.pdf>

を作成した。センター試験までの限られた時間で本書を活用して頂ければ幸いである。