

# 英語子音の使用頻度統計 に関する一考察

加藤力也

## 1. 課題の設定

### 1.1 使用頻度について

前回に発表した『英語子音の音韻機能について』<sup>1)</sup>において、ある与えられた形式 (form) の中で、それぞれの音素が、その形式の意味を伝えるために、対等の資格で協同して果す役割が積極的音韻機能と呼ばれることを述べた。そして各音素の積極的音韻機能の総和を、その音素の積極的機能負担量 <Positive Functional Burdening> と呼んだのであったが<sup>2)</sup>、これを統計的に求めたものが、各音素の使用頻度である。更にこれは、一言語の全ての材料を対象とすることは不可能であるため、一定の材料中での相対的な使用頻度 <Relative Frequency> として扱われる。

### 1.2 統計の類別

その統計は音素目録と一定の材料を決定しさえすれば、原理的には比較的簡単に行なわれ得る。しかし、音素目録を決定して実際に集計を行う際に、何を材料とするか、またその材料にどのような制限を加えるかによって、数値は勿論のこと、統計全体としての性質も非常に違ったものとなる。

統計の性質を大きく異ならせるのは、実生活において用いられたものを材料とするか、辞典を材料とするかである。辞典を材料とした場合には語 (word) を単位として、その中に生起する音素の統計を求める方法が考えられるが、この際、各々の語はそれぞれ同一の頻度、つまり一回の頻度を持つも

のとされ、語自身の持つ頻度は無視される。これに反し、実生活において用いられた材料に基づく場合には、単位となる語自身の頻度がその中に含まれる音素の頻度を倍加させることになる。

辞典に基づく場合には辞典を定めて、それに記載された全ての語を材料とする、任意に一定の量の語を抽出する、あるいは一定の音節の語のみを抽出する等の方法が考えられる。更に、ある分野の語彙を代表するに足るだけの相当数の延べ使用語 <Running Word> の中に現われた異語 <Different Word> の使用頻度の統計—語彙統計 <Word Count>—により決定した、例えばソーンダイクなどの単語表 <Word List> の中から使用頻度の高い一定数の語を材料とする場合も、各語の使用頻度を同一のものとみなす時は、性質上、この辞典に基づく統計の中に属するものと考えられるであろう。

実生活の中で用いられた材料は、Rebecca E. Hayden が行っているように<sup>3)</sup>、書かれたものと話されたものとに分けることが出来るであろう。しかし、このような区別は先に述べた語彙統計に於ても問題とされる所であり、両者の違いを判然とさせ得るかどうかは疑問である。話されたものとしては普通の会話、電話による会話、ラジオ放送、講演等、書かれたものとしては子供の読物、教科書、小説、劇、論文などが材料とされるのであるが、会話と子供の読物、講演と論文の相違はつけにくく、また特に劇などは話し言葉の材料として使われることも多いので、両種を区別することは簡単に出来ない。しかし、いずれを材料とするにせよ、単語の頻度が音素の使用頻度に影響を与えることは共通であって数値の相違が生じても、これらの統計は根本的に同じ性質のものであると考えられる。

### 1.3 各統計間の相関

前回の論文では使用頻度、すなわち積極的機能負担量のみならず、機能負担量 <Functional Burdening または Functional Load> 及び消極的機能負担量 <Negative Functional Burdening> を求める関係上、辞典に頼らざるを得なかった<sup>4)</sup>。使用頻度のみを求めようとする場合には、服部四郎氏も述べているように<sup>5)</sup>、辞典に基づくのは本筋ではなく、実際に用いられたものを

材料とする方が適切である。この観点から出発して、私の行った辞典に基づく統計が他の同種の統計とどの程度に密接な関係を持つのか、更に実生活の材料に基づく統計とどの程度の開きがあるのか調べてみたいと考えた。今回の論文は、(1)辞典に基づく統計の間の相関、(2)実生活の材料に基づく統計の間の相関、更に、(3)辞典に基づく統計と実生活の材料に基づく統計の間の相関を求めて、今後の研究の覚え書きとするものである。

## 2. 資料について

前章で述べた各統計間の相関を述べるに当たって、最も望ましいのは、これらの統計の基礎となった各資料が、英語の一方言を同一の音素目録によって採録したものであることである。しかし、実際にはこの二つの条件をみたした資料に基づく統計を四種見出すのは容易なことではない。今回の研究の対象とする統計もその点で完全なものとは言えないと考えるのであるが、以下各統計を、中心となる統計 I-1 に視点を合わせながら、この二つの点に注意して概説する。

### 2.1 統計 I-1

辞典に基づく統計の第一は前回の論文の第四章で発表した英語子音24についての相対的使用頻度である。これは public school で教育を受けた英国南部の人々の家庭での日常会話に聞かれる発音 Received Pronunciation (RP) を国際音声学協会の Phonetic Alphabet によって表記した Daniel Jones 著 *Everyman's English Pronouncing Dictionary* (11版, Dent & Sons, 1956) より、D. Jones の音素体系をそのまま採用して、(1)単音節語、(2)典型的発音、(3)規則的屈折形の除外の三条件を適用して求めたものである<sup>6)</sup>。

### 2.2 統計 II-2

辞典に基づく統計の第二は市河三喜編『英語学辞典』（研究社、1940）の p. 400, Functional Burdening（機能負担量）の項に添附された英語子音頻度表である。同表は単形態素、2音節までの語に生ずる24の英語子音の使用頻度を実数により示したものである。この表に関して更に詳細な知識を得た

いと考えたが、同項は木坂千秋氏の担当したものであって、同氏が死亡されたため、そこに書かれた事以外は不明であるとの報告を研究社を通して受けた。したがって、他の文献からの引用か、あるいは氏自身が実際に行なわれたものであるのかも不幸にして知ることはできないのであるが、音素目録は D. Jones のそれと一致している。また、/r/ が post-vocal に現われていないこと、/h/ が initial anti-consonant に 6 回しか現われていないのは、/hj/ のみが取上げられて、アメリカ英語の /hw/ は除外されているものと考えられるなどの点から、イギリス英語を対象としたものであろうと推察される<sup>7)</sup>。

### 2.3 統計 II-1

実生活の資料に基づく統計の第一は Godfrey Dewey 著 *Relative Frequency of English Speech Sounds* (Harvard University Press, 1923), p. 125 に記載されている英語音の使用頻度表である。この書の研究の目的は『今日書かれ話されている良い英語 (good English) にみられる、英語 (the English language) の種々の単音及び比較的普通な音結合すなわち音節と語の相対的出現頻度を決定すること』<sup>8)</sup> であった。この調査に使用された資料は新聞論説英語、ニュース英語、小説・短編小説・劇作を含む仮作物語、現代アメリカの演説、私信及び商業通信、現代の広告、宗教関係英語、科学英語、その他雑誌などの英語であって、これらから総計 100,000 の延べ使用語を抽出した。

資料にははっきり英国人の手になるものもいづらか含まれているが、抽出された単語を音声表記する際に、Revised Scientific Alphabet を使用している Standard Dictionary (Funk & Wagnalls, New York) の発音を適用しているので<sup>9)</sup>、発音に関する限りでは、方言性は無視されて、いわゆる General American になっているものと考えられる<sup>10)</sup>。

先に述べた英語音の使用頻度は、第 2 部の比較的頻度の高い語 (Particular Words と呼ばれる)、第 3 部の比較的使用頻度の高い音節に続き、第 4 部に取上げられているが、子音の総計は 62.10 %、母音は 35.32 %、二重母音は 2.58 % となっている。この表の中から子音に関する統計のみは Leonard

Bloomfield によって、英語に対する最近の子音音素の使用頻度として *Language* (Henry Holt, 1933) に採用されており<sup>11)</sup>、英国版 (George Allen & Unwin, 1935) では D. Jones 式の発音記号を以って示されている<sup>12)</sup>。本研究においては、記号統一の関係で、この表記を利用する。

G. Dewey の原典の表記と D. Jones のそれを較べると、前者が一般アメリカ音の表記であるため、母音の後の /r/ を子音に含めていること及び /hw/ が2つの子音として扱われることは当然であるが、子音に関する限りでは、その相違は記号上のものとどまっている。ただし、/iū/ 及び /iu/ を二重母音とするため<sup>13)</sup>、pure, few などは D. Jones のように /pj/, /fj/ とはなっていない。

## 2.4 統計 II-2

実生活の資料に基づく統計の第2は Rebecca E. Hayden が *The Relative Frequency of Phonemes in General-American English* (Word, vol. 6, p. 217~233, 1950) において発表した英語音素統計 (phoneme count) である。この研究の目的は一般アメリカ英語 (General-American English) の代表的な標本における音素の出現頻度を分析することであったが、とり分け、先に上げた G. Dewey などの研究が、書かれた材料に基づく研究であったこと、更に話された材料に基づくものであっても、一つの方言を扱ったものでないことを不満としてなされた統計である<sup>14)</sup>。

この研究のために選ばれた一般アメリカ方言 (General-American dialect) の標本は、カルフォルニア大学の外国語学生のためのオリエンテーションと英語計画においてなされた6つの講演であって、合衆国における外国学生の問題、教育の問題、労働問題、政治問題、農業森林問題等に話題が分れるよう配慮された。

6人の講演者の英語は差異が殆どなく、単一の音素体系すなわち一般アメリカ方言によって分析できるものであった。これらの講演を録音し、主として Kenneth L. Pike (*Phonemics*, p. 45, University of Michigan, 1947) の音素体系に従って転写がなされた<sup>15)</sup>。但し、K. L. Pike の成節子音 /m, n

### 英語子音の使用頻度統計に関する一考察

l, r/ を無視したこと, また子音 /y/<sup>16)</sup> は /hy/ (D. Jones による/hj/) と分析したことにより, D. Jones の表記により近いものとなっている。D. Jones の表記との相違点は, 一般アメリカ音であるため, 母音の後の /r/ を子音とすること, 及び 'where' と 'wear' の語頭の子音を区別することであるが, L. E. Hayden は K. L. Pike に従って, 'where' の語頭音を /w/ の1子音として取扱っている。D. Jones に従って記号を統一するとすれば, /w/ を /hw/ と解釈し直して, /h/ 及び /w/ の統計に組み入れなければならないのであるが, 煩雑さを避けるために, /w/ の統計を単に /w/ のみに加えることにした。その他の相違は記号上のものである。/w/ を含む25の子音の総計は全体の62.6%, 母音は37.4%となっている。

## 2.5 各統計の一覧表

第一表に示す通りである。

第 1 表

	S <sub>I-1</sub>		S <sub>I-2</sub>		S <sub>II-1</sub>		S <sub>II-2</sub>	
	頻出数	百分率	頻出数	百分率	頻出数	百分率	頻出数	百分率
p	582	5.96%	947	6.34%	7,620	3.29%	1,468	3.60%
b	429	4.39	754	5.04	6,750	2.917	1,075	2.64
t	939	9.61	1,687	11.29	26,550	11.47	4,945	12.14
d	519	5.31	830	5.55	16,050	6.94	2,091	5.13
tʃ	198	2.03	241	1.61	1,930	0.83	343	0.84
dʒ	179	1.83	272	1.82	1,630	0.70	325	0.80
k	883	9.04	1,271	8.50	10,100	4.36	1,942	4.77
g	334	3.42	498	3.33	2,760	1.19	745	1.83
f	389	3.98	407	2.72	6,860	2.96	1,051	2.58
v	153	1.57	298	1.99	8,490	3.67	1,518	3.73
θ	140	1.43	117	0.78	1,380	0.60	286	0.70
ð	43	0.44	64	0.43	12,780	5.52	2,180	5.35
s	1,111	11.38	1,427	9.55	16,970	7.33	3,186	7.82
z	235	2.41	207	1.38	11,070	4.78	1,539	3.78
ʃ	203	2.08	283	1.89	3,052	1.32	565	1.39
ʒ	6	0.06	22	0.15	170	0.07	19	0.05
m	448	4.59	797	5.33	10,360	4.48	1,870	4.59
n	679	6.95	1,143	7.65	26,980	11.66	5,179	12.71
ŋ	130	1.33	204	1.36	3,590	1.55	522	1.28
l	899	9.20	1,546	10.34	13,930	6.02	2,377	5.83
r	702	7.19	1,088	7.28	25,620	11.07	4,625	11.35
j	102	1.04	261	1.75	2,230	0.96	780	1.91
h	166	1.70	240	1.61	6,755	2.919	721	1.77
w	298	3.05	342	2.29	7,760	3.35	1,392	3.42
計	9,767	100.0	14,946	100.0	231,387	100.0	40,744	100.0

## 3. 手 順

§1.3 で設定した各統計間の相関を求めるために、ピアソンの係数を用いた<sup>1)</sup>。相関係数  $r$  を求める式は下記の通りである。

$$r = \frac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$$

$x, y$  はそれぞれ標本  $X, Y$  の各平均からの偏差であり、 $n$  は標本の大きさをなわち測定された対の数である。 $r$  の計算式は次のとおり。

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}}$$

但し、 $-1 \leq r \leq 1$

なお、順位相関的な見方は各段階の差が均一でないため、統計的な妥当性はうすいのであるが、各統計間の相関の内容的な差を示すのに役立つと考えられるので、ピアソンの係数を求めた後、補足として簡単なスケッチを試みることにする。

## 3.1 統計 I-1 と I-2 の相関

$$r = \frac{24 \times 662.3601 - 99.99 \times 99.98}{\sqrt{24 \times 657.1657 - 99.99^2} \sqrt{24 \times 681.5819 - 99.98^2}}$$

$$r = 0.9734 \dots \dots$$

$$r \approx 0.97$$

なお、第1.1図に示すように、両統計における各音素の順位を見ると、/s, t, l/ が第1位から3位までを占め、/k/ は同順位の4位。続いて、/r, n/ が5位と6位を占め、/p/, /d/, /m/, /b/ は同順位で7位から10位まで、/f, g/ が11位と12位を占め、更に /w/ が同順位で13位となっている。以下 /z, ʃ, tʃ, dʒ, h, v, θ, ŋ, j/ は /ʃ/ と /ŋ/ のみが同順位で14位から22位までを占め、最後に /ð/ と /z/ が同順位で23位と24位を占めている。14位から22位までの順位は乱れているが、各音素の頻度が小さいため両統計の相関値にはあまり影響を与えない。第13位までの頻度の合計は  $S_{I-1}$  で84.07%、 $S_{I-2}$  で85.21%である。



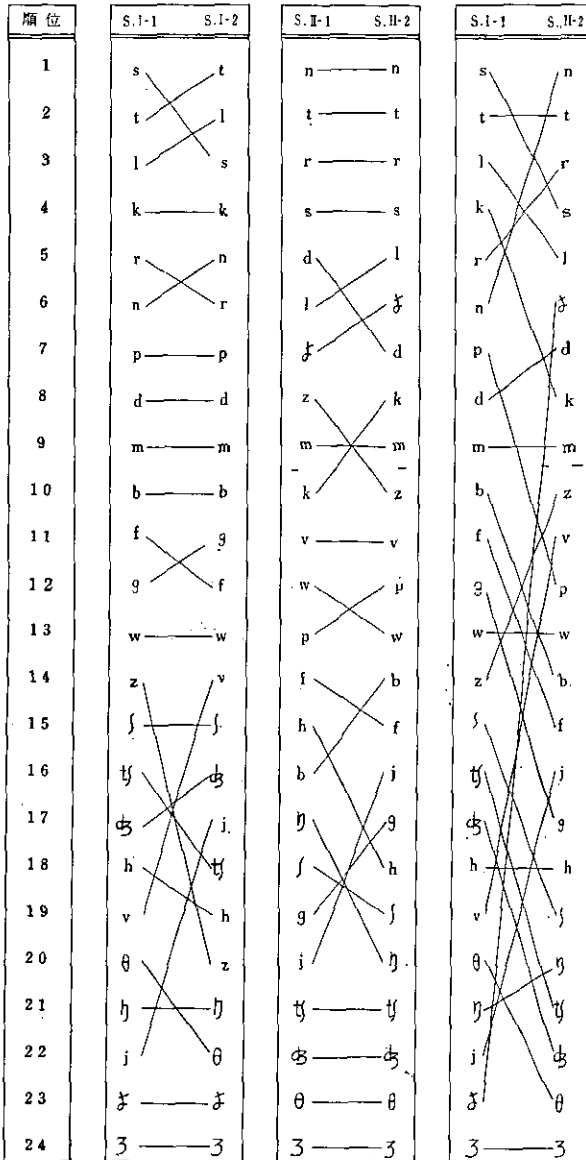
英語子音の使用頻度統計に関する一考察

第 1 図

1.1 図

1.2 図

1.3 図



## 3.2 統計 II-1 と II-2 の相関

$$r \approx 0.98$$

第1.2図に示すように、両統計における各音素の順位は、/n/, /t/, /r/, /s/ が同順位で1位から4位まで占める。続いて /d, l, ð/ が5位から7位まで、/z, m, k/ は /m/ のみが同順位で8位から10位までを占め、/v/ が同順位の11位で、更に /w, p/ が12位と13位を占めている。ここまでの合計は  $S_{II-1}$  で83.94%  $S_{II-2}$  で84.22%である。以下位14から20位までは /f, h, b, ŋ, s, g, j/ が入り乱れて順位を占め<sup>15)</sup>、最後の /tʃ/, /dʒ/, /θ/, /ʒ/ は順位が一致している。

## 3.3 統計 I-1 と II-2 の相関

$$r \approx 0.73$$

第1.3図に示すように、両統計における各音素の順位は、非常に入り乱れており、高順位の /t/ が同順位であることを指摘できるくらいで、到底前者のように、簡単な順位による記述は出来ない。§1.2で述べた如く、この順位の相違は、/r/ を除けば<sup>16)</sup>、主として単位となる語自身の使用頻度の影響であると考えられるが、個々の音素についてのこの問題の解明は現段階では不可能である。ただ、G. Dewey の統計表によると、17段階をへだたる /ð/ (Dewey のものとは16段階) については、その  $\frac{\text{Occurrence}}{\text{Items}}$  は他の子音音素が最高 /w/ の13.98であるのに対し、110.17となっていることを指摘するにとめたい。

## 4. 結 論

統計 I-1 と統計 I-2 は共に辞典に基づく統計であるが、その主たる相違は一方が一音節語までの資料、他が二音節語までの資料に基づいていることである。しかし、両者の間の相関度はきわめて高く、音節の数を増すことによって資料を増しても同じような使用頻度の統計の得られることが予想されるであろう。

統計 II-1 と統計 II-2 は共に一般アメリカ英語の実生活の資料に基づく統計であるが、その相違は R. E. Hayden に従えば、一方は書かれたものを資料とし、他方は話されたものを資料にしていることである。前にも指摘したように、話されたものと書かれたものとを区別すること自体が決して容易なことではないのであるが、とにかく両統計間の相関度は先の辞典に基づく統計間の相関度よりも更に高い。統計的な解釈によれば、殆ど完全に一致していると言ってよいほどである。したがって、子音の使用頻度に関する限りでは、書かれたものによる統計と話されたものによる統計の差を重視する必要はないように思われる<sup>20)</sup>。

最後に、統計 I-1 と統計 II-2 は、一方は辞典に基づく統計、他方は実生活の資料に基づく統計の代表として取り上げたものである。両者の相関度はピアソン係数0.73であり、これは統計的解釈にしたがえば、高い相関を示していることになる。しかし、前の二つの相関度に較べれば、はるかに低く、両者の質的な相違がはっきり現われており、どちらか一方をもって、他方の代りとするのは危険であろう。

#### 註

- 1) 『北海道英語英文学』第8巻(日本英文学会北海道支部, 1962)横書P.9~20。
- 2) 前掲論文 P.9 (§ 1.2)。
- 3) § 2.4 参照。
- 4) 前掲論文 P.11 (§ 1.3)。
- 5) 服部四郎『言語学の方法』(岩波書店, 1960), P.300。
- 6) 詳細については前掲論文 P.11~12(2章及び3章)を参照されたい。
- 7) しかし, final post-vocal における /j/, 及び final post-consonant における /l, n, (m)/ については疑問が残る。
- 8) 同書 P.3。
- 9) Ibid, p. 11。
- 10) Isaac, K. Funk: *New Standard Dictionary of the English Language* (Funk & Wagnalls Co., 1962) p. xii~xiii 及び xxi 以下参照。
- 11) 同書 P. 137。但し /h/ と /b/ の順位は間違っている。
- 12) 英国版 p. viii 参照。なお /g/ と /g/ については三宅・日野訳『言語』(大修館, 1962) p. xii 訳出要綱付記参照。

- 13) G. Dewey の前掲書 p. 136~138 参照。
- 14) 同論文 p. 217~218 (2章) 参照。
- 15) 前掲論文 p. 219 (脚註) 参照。
- 16) /y u ʃ/ 'huge'。K. L. Pike の前掲書 p. 45。
- 17) 岩原信九郎『教育と心理のための推計学』(日本文化科学社, 1957) p. 128。
- 18) G. Dewey (S<sub>II-1</sub>) の /j/ は /iu/ 60回 /iū/ 1080回を加えると第18位に, また R. E. Hayden (S<sub>II-2</sub>) の /h/ は /w/ 241回を加えると第16位に上るため, /f, h, b/ が14位から16位まで, /ŋ, j, ʃ, g/ が17位から20位までと更にまとまったものになる。なお S<sub>II-2</sub> の /w/ は /w/ を加えなくても同じ順位である。
- 19) S<sub>II-2</sub> の /w/ は where などの頭音を含めてある。
- 20) 母音に関してはこの限りではない。辞典に基づく場合もそうであるが, 書かれたものを表記する際には strong form を採用するのが普通であるが, 話されたものを実際に録音すれば weak form が多く現れ, したがって, 弱音節に用いられる母音の頻度が大きくなるであろう。