

国立国語研究所学術情報リポジトリ

表記のゆれを測る

メタデータ	言語: Japanese 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 佐竹, 秀雄 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002873

表記のゆれを測る

佐竹 秀雄

1. ねらい

日本語の表記には、ひらがな・カタカナ・漢字・ローマ字・数字などが用いられる。これらの文字は互いに独立するものではなく、関連をもっている。したがって、日本語で一つのことばを書き表すとき、表記の形態は唯一ではない。かと言って、まったく自由というわけでもない。表記の選択にはさまざまな要因が働き、その結果、表記はある場合には固定され、ある場合には大きくゆれるということになるのであろう。日本語の表記法を明らかにするためには、表記のゆれの実態をとらえ、さらにそのゆれを生み出す要因をさぐらねばならない。

そこで、まず表記のゆれの実態を把握するための尺度について考察してみた。今回はその一つの試みの発表である。

2. 表記のゆれのパタン

表記のゆれと一口に言っても、その内容はさまざまである。例えば、かなづかいにおける新旧のゆれもあれば、ひらがなで書くかカタカナで書くかというゆれもある。さらには、縦書きにするか横書きにするか、手書きか印刷かといった表記行動に属するようなゆれも考えられうる。しかし、ここで取り上げるのは、上述の立場から、語の表記形式のゆれである。すなわち、語がどのような表記形式で書き表されるかということである。

語の表記形式のゆれとしては、次のような対立のパタンが見られる。

I 同一文字体系内での対立によるゆれ

- (A) 二種以上の異なる漢字の対立 例(尚-猶)
- (B) 二種以上の異なる字体の文字の対立 (學-学) (峰-峯)
- (C) 二種以上の送りがなの対立 (行なう-行う)
- (D) かなづかい・かな用法上の対立 (ついに-つひに)
- (E) カタカナの外来語表記上の対立 (コンベア-コンベヤ)
- (F) ローマ字のヘボン式表記と日本式表記の対立 (Fuji - Huzi)
- (G) 数字で桁の単位の有無の対立 (一二三四五-一万二千三百四十五)

II 異なる文字体系間の対立によるゆれ

漢字・ひらがな・カタカナ・ローマ字・数字などの間での相互の対立

III その他

- (H) くり返し記号と文字との対立 (人々-人人)
- (I) その他の記号類と文字との対立 (%-パーセント)

ここで注意すべきことがある。その一つは、何と何を同じ文字あるいは同じ語と認め、何と何を異なる文字・異なる語と認めるかということである。その認定の仕方によっては、上の対立のいくつかはゆれとは認められなくなる。例えば字体の対立は無視してよいとの立場に立てば(B)は不必要である。同じと認めるか、異なると認めるかの考え方が違えば、それに応じてゆれの問題も違ってくる。よって、同一文字体系間の対立は異なっているとは認めないとか、ひらがな・カタカナは合わせてかなとして扱うという立場もないとは言えない。

そして、もう一つ注意すべきことは、実際の表記においては、上述の種々のゆれが複合された形で出現することである。例えば「現代雑誌九十種の用字調査」^{注1)}では、副詞の「きっと」の表記として

きつと(28例)・きつと(28例)・キツト(2例)・屹度(1例)・きッと(1例)・キツト(1例)

が見られる。「きつと」「キツト」などは特殊な表記形式だと言えよう。

3. 表記のゆれの尺度

それでは、このような表記のゆれをいかにすれば測ることができるか。

まず比喩的なモデルを想定してみよう。「100個のボールを10個の箱に好きなように入れる」ことにする。ボールとは一つの語であり、一つの箱は一つの表記形式を表す。とすると、表記のゆれの度合いとは、ボールの入った箱の数とその中のボールの数に、どれほど散らばりがあるかということになる。ボールがより多くの箱に、より平均的に入っているほど、散らばりが大きいと言えよう。

こう考えれば、表記のゆれを測るには、一種の散らばり度を測定すればよいことになる。そこで散らばりの尺度を示す分散の考え方を応用してみる。分散の基本的な考え方は、 $S = \frac{1}{N} \sum (y_i - \bar{y})^2$ で示されるように、実測値に対して何らかの意味で基準となる理論値を求め、それと実測値とのへだたりの平方和によって算出するというものである。

いま、ある語が N 回出現したとする。その表記形式は G_1, G_2, \dots, G_r の r 通りあり、各表記形式の出現回数は L_1, L_2, \dots, L_r だとする。当然 $N = \sum L_i$ であり、各表記形式の出現率は $P_i = L_i / N$ で求められ、 $\sum P_i = 1$ となる(表1参照)。そして分散をこの出現率で考えると、

$$C = \sum (P_i - \bar{p})^2 \dots \dots \dots \textcircled{1}$$

と表せる。

次にこの \bar{p} なる理論値だが、これは P_i の平均値では意味がない。なぜなら、表記形式のあり方は語によってさまざまであり、例えば常に漢字表記されるものと、まったく漢字表記されないものがあり、それらを一緒にして平均値をとるような一義的な算出法は無意味だからである。

そこで、表記のゆれが最大になるという極限状態を想定してみよう。これは表記形式 r が無限にあり、それらの各表記形式の出現率が平等な状態、すなわち P_i がほとんどゼロに等しい状態である。この状態を基準にとり、理論値 $\bar{p} = 0$ とするわけである。したがって、これによって求められるのは、表記のゆれのもっとも大きな状態からのへだたりである。ゆれが最大の状態からのへだたりが大きいことは、実際にはゆれがもっとも小さいことを意味する。よって $\bar{p} = 0$ を $\textcircled{1}$ に代入して、 $C' = \sum P_i^2$ を求めるとは、語の表記形式の集中度を測ることに等しい。そこで再びゆれの大きさを示すものとして、 C'

の最大値1からC'を差し引いたものをとることとする。つまり、表記のゆれの尺度Sを

$$S = 1 - \sum P_i^2 \dots\dots \textcircled{2}$$

と定めるわけである。

さきあげた「きつと28/きつと28/キツと2/屹度1/きつと1/キツと1」の例で演算してみると、

$$S = 1 - (0.459^2 + 0.459^2 + 0.033^2 + 0.016^2 + 0.16^2 + 0.16^2) = 0.577$$

となる。

4. ゆれの尺度の適用と反省

以上で、この論の主たる目的は一応達せられたわけであるが、このゆれの尺度を具体的なデータに適用すればどうなるだろうか。「新聞の語彙調査」^{注2)}のデータ(昭和41年の新聞)の一部に対して測定してみた。対象にした語は、使用度数20以上の形容詞(45語)、形容動詞語幹(42語)、副詞(67語)及び一部の動詞(80語)である。^{注3)}品詞及び表記の確認は、KWIC^{注4)}を見ながら行った。また表記形式のゆれとして認めたのは、異なる文字体系間の対立と二種以上の漢字表記の対立によるものである。

結果の一部を表2、表3とグラフに示す。

これらからどのようなことが考えられるのだろうか。尺度の有効性や調査の方法の点で、まだ問題が残るので、断定的な態度はさげねばならない。よって、以下目につくことをいくつか推測として述べる。

- A) 表2から動詞のゆれが他より大きいことがわかる。これは、動詞が形容詞や形容動詞語幹に比べて、一つの語が含みもつ意味が多様であることと関係があるのではないか。つまり意味の差異を表記の違いで示している可能性があると思われる。
- B) 表2から形容詞の「一表記あたりのゆれ」が他に比べて小さいことが認められる。これは使用度数の高い語のゆれが小さいためである。つまり、形容詞では使用率の高い語は表記が安定していると言えよう。
- C) グラフによれば、形容動詞語幹でゆれないもののがかなり多い。このほとんどは漢語であり、漢字表記である。それに対し、表3のゆれの大きい形容動詞語幹の7語のうち、和語が6語を占めている(調査対象42語のうち和語は12語)。これは漢語の表記が和語に比べて安定していることを示すものと考えられる。
- D) 表3で、ゆれの大きな語の中に、「面白い・素敵・真面目」のような当て字・熟字訓の表記を用いる語が見られる。また「固い・主な・盛んな」のような、昭和41年当時は当用漢字音訓表で認められていない表記が用いられている語がある。この事実は、当用漢字表・当用漢字音訓表による制約が表記に与えた影響が働いたと思われる。

以上、表記のゆれと、品詞・使用率・語種・表記上の制約とのかかわりをかいま見たわけであるが、この他に書き手や書く場面といった、さらに大きな要因があろう。それらの要因を的確に分析するためには、今の尺度では不十分かもしれない。今のままでは、使用度数が低い場合、ゆれの誤差が大きくなる欠点があり、改良の余地がある。また統計学的な検討も必要である。これらは今後の課題としたい。

- 注1) 国立国語研究所による「現代雑誌九十種の用語用字」の調査研究（データは昭和31年の雑誌）
 注2) 国立国語研究所報告38『電子計算機による新聞の語彙調査（Ⅱ）』（昭和45年）
 注3) 品詞別度数順短単位表をもとに、度数20以上のほとんどすべての形容詞・形容動詞語幹・副詞を調査。動詞については度数20以上の約40%の語を調べた。
 注4) 電子技術総合研究所の植村俊亮氏によって、マイクロフィッシュ化されたもの。

表1 表記のモデル

表記形式	G_1 G_2 …… G_i …… G_r	計
出現度数	L_1 L_2 …… L_i …… L_r	N
出現率	P_1 P_2 …… P_i …… P_r	1

表2 品詞別にみた表記のゆれ

	対象語数	延べ語数	一語あたりのゆれ	一表記あたりのゆれ
動 詞	80	10151	0.218	0.177
形 容 詞	45	4637	0.123	0.068
形容動詞語幹	42	1412	0.119	0.107
副 詞	67	4624	0.132	0.137

表3 ゆれの大きな語 (S>0.3)

《形容詞》		《副詞》	
1. 堅い	.578	1. いっそう	.522
2. くわしい	.482	2. はじめて	.515
3. おもしろい	.477	3. まったく	.499
4. 忙しい	.463	4. 直ちに	.498
5. 楽しい	.454	5. つねに	.497
6. 激しい	.361	6. さきに	.458
7. すばらしい	.349	7. 最も	.454
《形容動詞語幹》		8. 果たして	.453
1. すてき	.550	9. 次第に	.444
2. まじめ	.540	10. 一番	.432
3. 主	.500	11. とくに	.429
4. 大変	.493	12. なお	.404
5. たしか	.480	13. ときどき	.326
5. 盛ん	.480	14. たとえば	.324
7. 豊か	.305		

表記のゆれの大きさによる語の分布

