N-005　線形回帰分析による「漢字の将来」の予測（1963）と 50 年後の漢字含有率の実際
～分析方法と結果の再検討，および統計教育への教訓～

A Future Prediction of the Usage Rate of Chinese Characters in the Japanese Language in 1963 by Linear Regression Analysis, and Their Actual Usage Rate after Fifty Years: A Review of the Methods and Results, and A Lesson on Statistical Education

綾　皓二郎†
AYA Kohjiro

１．はじめに

今から五十年前に，当時若手の言語心理学者であった安本美典は，雑誌『言語生活』(1963.2）に「漢字の将来―漢字の余命はあと二百三十年か―」と題する論文を発表した(1)。その内容は，統計的な推測によれば，小説中の漢字の含有率は，西暦 2190 年頃にゼロになるというものであった。この論文は，日本語学者の野村雅昭が日本語の将来を論ずるときの基本文献としているように(2)，その後の漢字論・日本語表記論に大きな影響を及ぼした。本報告では，この論文における安本の回帰分析を再検討し，彼の導いた結論の問題点を指摘する。さらに，戦後の出版物における漢字含有率の推移を宮島達夫の論文(3) などを参照して考察する。

２．安本美典の分析と結果

2.1 安本美典の調査方法

小説の文中に出現する漢字含有率を調べ，それをもとに統計的推計を行った。1900 年から 1954 年までに発表された，100 人の作家による 100 編の小説から各 1000 字を抽出して，その漢字含有率を調べた。小説は，筑摩書房刊の『現代日本文学全集』から一人の作家につき一作品を選び出している。調査年次は，欠けている年もあれば，最大 6 編を調べている年もあるが，年次ごとの漢字の数が表として掲載されている（安本第 1 表)。(1)

2.2 安本美典の分析結果

原データを 5 年ごとに区切って，漢字含有率の平均値を求めたもの（安本第 1 図）と原データ全体を対象にして，回帰分析したもの（安本第 2，3，4 図）がある。以下では，まず彼の回帰分析の結果を要約する。説明変数 x を「西暦年数」，目的変数 y を「千字中の漢字の数」として，最小自乗法により回帰式を求めている。

・線形近似：　$y = -1.241\,x + 2726.17$　　（1）　(安本第 2 図)
・指数近似：　$y = 360.17 \exp(-0.0038659\,x + 7.34521)$　　（2）　(安本第 3 図)
・$dy/dx = -a/y$ ：　$y\text{^}2 = -1074.14\,x + 2181603.11$　　（3）　(安本第 4 図)

安本は，漢字の使用の度合は，ほぼ直線的に減少している，将来も直線的に減少すると考える，という二つの仮定をおいて，（1）式を求めた。（1）式によれば，50 年ごとに約 62 字ずつ減少していき，2000 年には漢字の含有率は 23.8%（小学校 5 年生の国語の教科書と同程度)，2191 年中には，漢字の含有率はゼロとなる。

（2）式は，（1）式に対する，一つの補正である。漢字が少なくなればなるほど，漢字保護の運動が高まり，漢字の減少の勾配は緩やかになる可能性がある。漢字減少の勾配　dy/dx は，そのときに用いられている漢字の使用度 y に比例して小さくなると考える。この場合には，$dy/dx = -by$ の微分方程式が成り立つ。未定定数を最小自乗法で求めれば，これは指数近似を適用したことになる。（2）式によれば，2000 年で 24.5%，2600 年で 2.4%の漢字含有率となる。

（3）式は，（1）式に対する，もう一つの補正である。漢字が少なくなればなるほど，漢字を減少させようとする勢力が力を得て，漢字の減少の勾配は急速になる。漢字減少の勾配は，そのときに用いられている漢字の使用度に反比例して増大すると考える。この場合には 2000 年で漢字の含有率は 18.3%，2031 年でゼロとなる。

安本は（2）（3）は両極端であり，当時の段階では，漢字含有率は（1）式で表されるようにほぼ直線的に減少していくと結論づけた。

図 1 は，著者が安本の第 2，3，4 図を原データに基づいて作成して一つの図にまとめたものである。

† 石巻専修大学理工学部

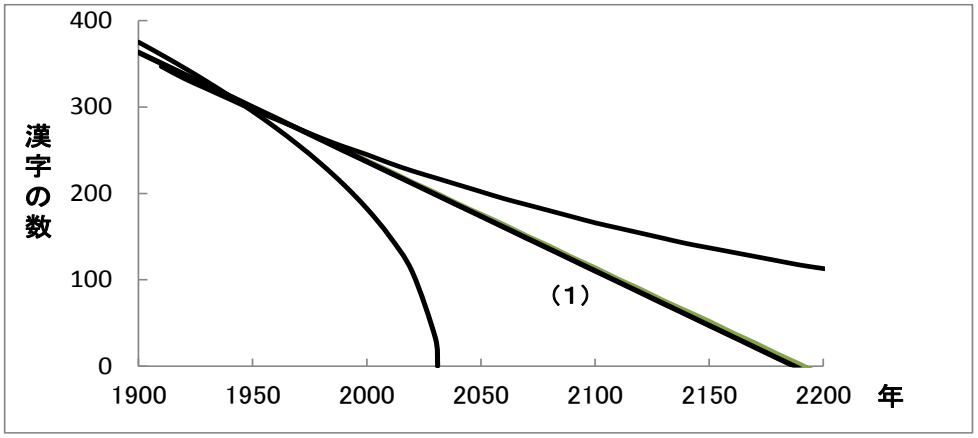

図 1　安本の論文の第 1,2,3 図を原データに基づいてまとめた図

図 1 で，（1）が回帰直線を表す。ただし，回帰式は，$y = -1.264\,x + 2765.4$ であり，安本の式と比べると，回帰係数と切片の値が若干異なった値が得られた。そのため，漢字含有率がゼロとなるのは，2187 年となる。

３．宮島達夫の分析と結果

3.1 宮島達夫の調査方法

国立国語研究所の日本語学の室長である宮島達夫は，安本による漢字の将来の推定が正しかったどうかを調べた結果を雑誌『言語生活』(1988.3）に「『漢字の将来』その後」と題する論文で発表した。(3)

調査対象は安本と同様に小説とし，芥川賞受賞作品（『文藝春秋版』）を選んだ。1935 年から 1985 年までに受賞した 94 人の作家による 94 編の小説から各 1000 字を抽出して，その漢字含有率を調べた。ただし，1945 年から 1948 年は受賞作品がない。また，該当作が 1 作しかない年もあれば，4 作ある年もある。原データは，年次・受賞作・字種ごとの字数からなる表として掲載されている（宮島の原表）

3.2 宮島達夫の分析結果

漢字含有率を 5 年ごとに区切って平均した結果の表を図にしている（宮島図 1)。図 2 は，著者が宮島の表 1 に基づいて作成した，宮島の図 1 である（雑誌と新聞のグラフは除く)。宮島の分析では，20 世紀前半に漢字が減る傾向を見せた点については，安本と結果が一致したが，漢字の含有率が減ったのは 1960 年までで，それ以後は減っていない（ほぼ横ばい)，という結論を得た。これは，1946 年の「当用漢字」(1850 字）から 1981 年の「常用漢字」(1945 字）への変化に象徴される，言語政策の保守化に対応する，としている。

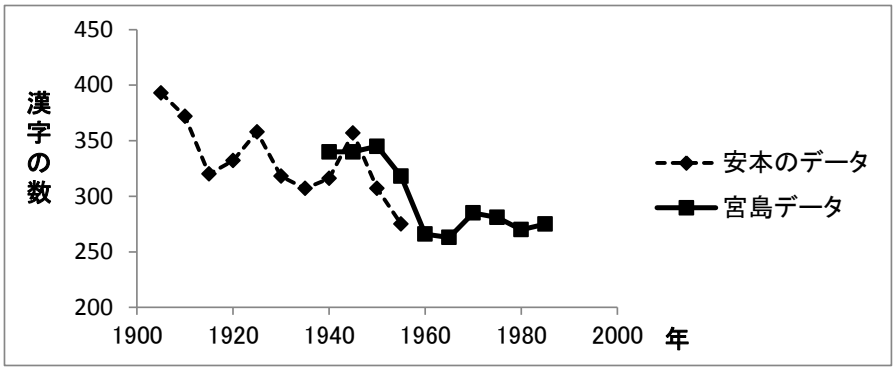

図 2　宮島の図 1

４．安本の分析の問題点と結果の検討

安本の論文の分析で，まず問題となることは，原データの散布図および相関係数，決定係数（単回帰の場合は相関係数の 2 乗)，分散分析の結果が載せられていないことである。これらは今では表計算ソフトを使えば容易に求まる。散布図と回帰直線を図 3 に示す。相関係数は −0.377 で，やや相関がある程度であり，決定係数は

0.114 である（回帰式 $y = -1.264\,x + 2765.4$ ）。安本のデータを分散分析した結果を表 1 に示す。

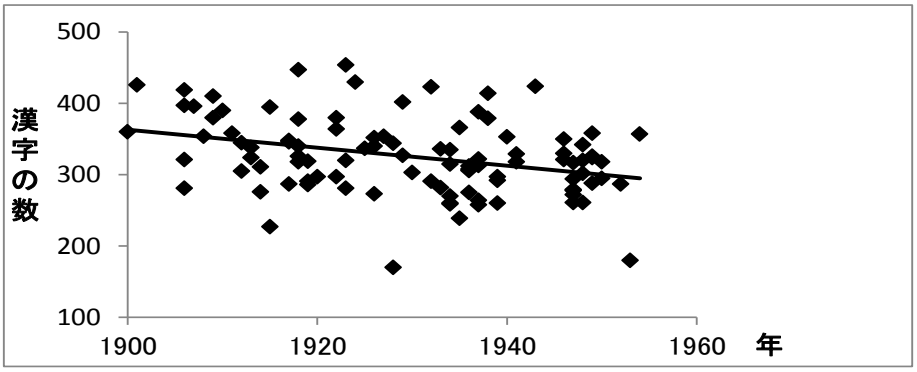

図 3　安本のデータの散布図と回帰直線

表 1　安本のデータの分散分析（Excel による）

回帰統計						
重相関 R	0.337005					
重決定 R2	0.1135724					
補正 R2	0.1045272					
標準誤差	50.572526					
観測数	100					
分散分析表						
	自由度	変動	分散	分散比	有意 F	
回帰	1	32113.282	32113.282	12.556118	0.0006068	
残差	98	250642.88	2557.5804			
合計	99	282756.16				
	係数	標準誤差	t	P−値	下限 95%	上限 95%
切片	2765.3779	688.51426	4.0164424	0.0001159	1399.0437	4131.712
X 値 1	−1.26432	0.3568038	−3.543461	0.0006068	−1.972386	−0.556255

表 1 によれば，確かに回帰係数の傾きは有意ではあるが，決定係数は 0.114 であるから，図 3 に見るように，回帰直線の当てはまりの程度はきわめて悪い（線形回帰は説明モデルとして弱すぎる）。安本は，散布図や決定係数を示すことなく，さらに回帰直線を原データの調査年間の 5 倍も 2200 年頃まで外挿して，小説中の漢字の含有率を予測している。決定係数が 0.5 以下のより小さい場合は，分散分析表の F 検定が棄却されたからといって，回帰式を予測に用いるのは要注意である。なぜなら，F 検定では単に回帰係数が 0 ではないといっているにすぎないからである。回帰係数が 0 でないからといって，予測に役立つわけではない。[4] さらに，線形回帰分析においては，回帰式はあくまで観測されたデータに基づいて求められたものであり，データが異なれば回帰式も当然異なったものになることに注意しなければならない。したがって，因果関係を論ずることなく，原観測データ範囲を大幅に超える説明変数を用いた，外挿による回帰予測は，信頼できない結果を生ずることに注意すべきである。図 3 の散布図と決定係数の値 0.114 が論文に載せられていたら，多くの人は安本の大胆な外挿に疑問をもったと思われる。

安本の論文は，『文章心理学の新領域』にも収録されている（回帰式と図には変更はない）。[5] その「用語のてびき」で【最小自乗法】【回帰】などを説明しているように，安本は計量言語学の専門家でもある。その彼がなぜこのような分析結果を発表したか，また 1963 年から 2001 年に至るまでの論文や著書[6] において，この問題に関する説明の補足や訂正は一切されていないなど，疑問が残る。さらに，国立国語研究所言語計量研究部長を務めた野村雅昭は『漢字の未来』という著書の中で，統計的分析方法を検討することなく，安本の結果をそのまま掲載して，「漢字はなくなるか」を論じている。[2] 安本の論文は，2009.12.7 の朝日新聞の記事[7] や 2011 年の田中克彦の著書[8] に見られるように，今でもかなりの影響力がある。

５．宮島のデータの回帰分析

宮島は，安本と違って，相関係数を求めている（安本のデータで −0.377，芥川賞の作品で −0.463）。しかし，回帰分析は行っていない。そこで，1935 年から 1985 年の原データに基づいて，線形単回帰分析を行うと，回帰式は $y = -1.711x + 3652.2$，決定係数は 0.214 である。このとき，安本と同様に外挿すれば，2135 年に漢字は消滅するという結果が得られる。他方，1955 年から 1985 年のデータを用いて，回帰直線と決定係数を求めると，$y = 0.112x + 54.84$，相関係数 0.02，決定係数 0.0004 となり，漢字含有率はランダムに散布していることがわかる（平均値は 27.6%）。また宮島は，1988 年の時点では，日本語ワープロの普及が漢字の使用にどのような効果をもつかを推定するのは困難である，と述べている（注：ワープロソフト「一太郎」の販売は 1985 年）。

６．戦後の出版物における漢字含有率の推移

- 新聞
  - 1955 年 朝日・毎日・読売 46.6%，1966 年 朝日・毎日・読売 38.7% (9)
  - 1971 年 共同通信 46.1%，(9) 1993 年 朝日 41.5% (10)
  - 2003 年 日経・朝日・読売社説 42.1%，2011 年 河北・朝日・読売社説 43.4% (11)
- 雑誌
  - 1963 年 週刊誌 週刊新潮 43.6% サンデー毎日 34.5% 女性自身 28.7% (9)
  - 1966 年 中央公論 37.8%，1976 年 中央公論 38.0% (3)
  - 1980 年 週刊誌 政経系 38.3%，新聞系 32.1%，出版系 32.0%，大衆系 29.5% (9)
  - 1991 年 若者向け雑誌 20% (12)
- 辞典：漢語の収録比率
  - 1956 年 例解国語辞典 53.6%，1969 年 角川国語辞典 52.9% (9)
  - 2002 年 新選国語辞典第 8 版 49.1%，2008 年 新選国語辞典第 9 版 49.4%

安本による予測では 2013 年の漢字含有率は 22%である。しかし，実際にはそれほどまでには減少していない。現在では漢字の含有率がおおよそ 30～40%であると読みやすいことが経験的に知られている。コンピュータと仮名漢字変換ソフトの開発と普及によって「漢字仮名交じり文」は，現代日本語の表記法として完全に定着し，今日の漢字含有率の安定（30～40%）をもたらしているといえる。(13)

７．おわりに

回帰分析法は実験や卒業研究でよく使われる統計的方法であるので，安本の分析は今日の統計・情報教育への有益な教訓を含んでいる。回帰分析では第一に散布図を描き，相関係数と決定係数を求めておく必要がある。

参考文献

(1) 安本美典：“漢字の将来―漢字の余命はあと二百三十年か―”言語生活，137 号，pp.46-54（1963）
(2) 野村雅昭：「漢字の未来 」pp.225-240，筑摩書房（1988）；新版 pp.195-208，三元社（2008）
(3) 宮島達男：“「漢字の将来」その後”，言語生活，436 号，pp.50-58（1988）
(4) 新村秀一：「JMP による統計レポート作成法」p.113，丸善(2007)
(5) 安本美典：「文章心理学の新領域」第 III 章，pp.140-174，誠信書房（1966）
(6) 安本美典：「説得の文章術」p.100，宝島社（1999，2001）
(7) 朝日新聞夕刊：“ニッポン人脈記 漢字の森深く：9 こんな文字，まっぴらや”，2009.12.7
(8) 田中克彦：「漢字が日本語をほろぼす」p.262，角川書店（2011）
(9) 宮島達男他：「図説日本語―グラフで見ることばの姿（角川小辞典）」角川書店（1982）
(10) 野崎浩成，清水康敬：“新聞における漢字頻度特性の分析と NIE のための漢字学習表の開発”，日本教育工学雑誌 24(2)，pp.121-132（2000）
(11) 千葉祐真：“1963 年における「漢字の将来」の予測と 2011 年における漢字使用率”，石巻専修大学理工学部卒業研究（2011）
(12) 佐竹秀雄,佐竹久仁子：「日本語を知る・磨くことばの表記の教科書」p.216，ベレ出版（2005）
(13) 綾 皓二郎：“再読 梅棹忠夫著『知的生産の技術』(1969)”，2012 PC Conference 論文集，pp.383-386（2012）