新聞における漢字使用の実態調査

横山詔一（国立国語研究所）・野崎浩成（名古屋市立大学）・米田純子（国立国語研究所）

１．目的

国民各層が社会動向を知る情報源として、我が国の新聞が果たしてきた役割はきわめて大きい。その意味で新聞の用字調査は、社会的に意義あるものだと言えよう。1995 年以後、大量の新聞記事を電子化した全文データベースが CD-ROM で市販されるようになり、新聞というジャンルの母集団にほぼ匹敵する規模のデータを誰でも入手可能となった。そこで、本研究は市販されている朝日新聞 CD-ROM をコーパス（corpus：言語分析の資料）として活用することにより、新聞メディアにおける漢字使用の状況を統計的に分析することを試みた。

新聞記事の漢字調査の代表は、国立国語研究所が 10 年の歳月をかけて公刊した『現代新聞の漢字』（1976）であろう。そこでは 1966 年の朝日・読売・毎日の新聞３紙 1 ケ年分の記事を対象に抽出比 60 分の 1 でランダム抽出して漢字調査を実施した。標本として得られた漢字の述べ字数は 991,375 字という大規模なもので、電子計算機による我が国で最初の漢字使用率基準表が作成された。その使用率の推定精度はきわめて高く、常用漢字表の制定や JIS 漢字の選定、さらには失語症診断テストの作成や文字認知実験など、幅広い学問領域において活用され、諸学界から高い評価を受けている。

しかし、その調査からすでに約 30 年も経過していることから、『現代新聞の漢字』（1976）の結果をそのまま現代の新聞記事にあてはめてよいものか疑問の声も聞かれる。近年の社会情勢の急激な変化やワープロ、パソコンの普及に伴うテキストデータ処理技術の発展によって、日本語における漢字使用の実態が 1966 年当時とは変化してきているのではないかという予測も立てられる。以下、この点を中心に計量国語学的な側面から検証を試みた。

２．方法

材料

＜1993 年コーパス＞

1993 年１月１日から同年 12 月 31 日までに発行された朝日新聞朝夕刊１ヶ年分の電子化された記事を用いた。このテキストデータは、CD-HIASK'93（朝日新聞社・紀伊国屋書店・日外アソシエーツ，1994）に格納されている新聞記事から、記事見出し部分を削除したものであった。漢字の延べ字数は 23,408,236 字、異なりは 4,476 字であった。

＜1966 年（昭和 41 年）コーパス＞

1966 年発行の朝日・毎日・読売３紙について、朝夕刊１ヶ年分から国立国語研究所が面積比 60 分の１でランダム抽出した記事に基づく。漢字の延べ字数は 991,375 字、異なりは 3,213 字であった。

手続き

漢字の定義は、JIS X0208（1990 年版）の区点コード表で第 16～47 区分に属する第１水準漢字集合、第 48～84 区分に属する第 2 水準漢字集合、両者合わせて 6,355 文字とした。ちなみに、JIS コードは 1978 年の制定以来、1983 年と 1990 年に大きな改訂・変更が加えられている。国立国語研究所では JIS C6226（1978 年版）で電子化データを保有しているため、異体字の字形を出来る限りチェックして JIS X0208 の区点コードに変換した。

３．結果と考察

1966 年と 1993 年の漢字使用頻度の相関を計算した。その結果、ｒ＝.95（ｎ＝4,535）に達し、全体的傾向として見た場合には時代変化の影響がさほど大きくはないことが明らかになった（図１）。特に、高頻度漢字は安定性が高い。

次に、常用漢字表以外の漢字（以下、表外字と呼ぶ）のうち、交ぜ書き解消のため新聞協会で 1981 年から使用が認められた 10 字について使用順位の時代変化を追った（表１）。

表１　使用順位の変動

漢字	区点コード	66年順位	93年順位	上昇順位数
亀	2121	1,289	1,354	-65
舷	2431	2,522	2,289	233
痕	2615	2,522	2,057	465
挫	2635	2,150	1,890	260
哨	3005	3,214	2,118	1,096
狙	3332	1,983	854	1,129
腫	2880	2,522	1,990	532
賢	3153	2,150	1,783	367
冤	4945	2,788	2,319	469
拉	5739	2,522	2,385	137

1966 年に比べて 1993 年の使用順位が上昇した漢字は「亀」を除く９字であり、２項分布に基づくサイン検定の結果、1993 年の方が有意に使用順位が上昇していた（$p<.05$）。

この結果は、表外字使用に関する新聞各社の姿勢が、紙面における漢字使用に現実に影響を及ぼしたことを裏付けるものであろう。また、表１のうち８字が JIS 第１水準に属することも見逃せない。記者がワープロで原稿を書くようになった今、JIS 漢字の枠組みが新聞メディアにも大きな影響を及ぼしつつあると考えられよう。

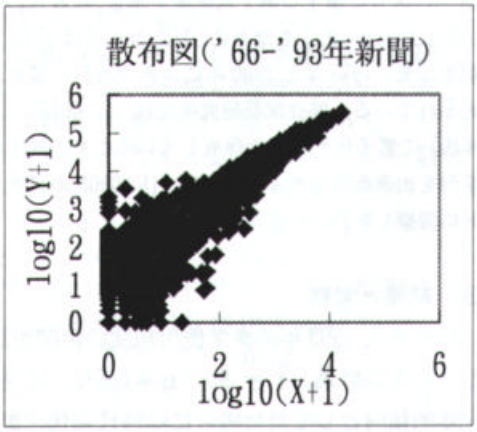

図１．対数変換後の散布図

ただしX：1966 年の文字頻度

Y：1993 年の文字頻度

４．参考文献

○学術論文（レフェリー付）

1. 野崎浩成・横山詔一・米田純子ほか（1997 年３月公刊予定）「文字使用に関する計量的研究－日本語教育支援の観点から－」『日本教育工学会論文誌』日本教育工学会

○口頭発表

1. 横山詔一・野崎浩成（9月）「コーパスを利用した日本語環境の分析」，日本行動計量学会第 24 回大会特別セッション，pp.138-139
2. 野崎浩成・横山詔一（9月）「新聞と雑誌における漢字使用頻度の分析 －心理学での材料統制の観点から－」，日本行動計量学会第 24 回大会，pp.266-267
3. 横山詔一・野崎浩成（9月）「心理学のための漢字頻度基準表の作成(1) －新聞コーパスを用いた時代変化のマクロ分析－」，日本心理学会第 60 回大会，p.599
4. 野崎浩成・横山詔一（9月）「心理学のための漢字頻度基準表の作成(2) －時代変化のミクロ分析と Internet による基準表の有効利用－」，日本心理学会第 60 回大会，p.600
5. 近藤公久・天野成章・横山詔一・野崎浩成（9月）「漢字の親密度と出現頻度の相関」，日本心理学会第 60 回大会，p.601
6. 横山詔一・野崎浩成（9月）「頻度辞書の公開を目指して」，日本心理学会第 60 回大会ワークショップ講演資料
7. 横山詔一・野崎浩成・米田純子（9月）「新聞の漢字使用順位に影響する要因の分析」，計量国語学会第 40 回大会，p.7
8. 野崎浩成・横山詔一ほか（9月）「漢字使用頻度の時代的変化に関する考察」，計量国語学会第 40 回大会，p.8
9. YOKOYAMA Shoichi（12 月）「Human Information Processing of Kanji Word」，Workshop at Monash University '96
10. NOZAKI Hironari, CHIKAMATSU Nobuko and YOKOYAMA Shoichi（12 月）「Japanese On-Line Resource Materials and the Japanese Character Frequency Database」，Workshop at Monash University '96

【謝辞】

本研究の遂行にあたり、重点領域研究『人文科学とコンピュータ』（課題番号：07207105、領域代表者：及川昭文）と国際学術研究（共同研究）『海外日本語学習リソース提供システムの実験研究』（課題番号：08044015、研究代表者：柳澤好昭）の援助も受けた。