資　　料

最近の日刊紙における文字使用状態について*

林原　初穂**　矢野　雅久**　黒崎　悦明**

1. まえがき

漢字を含む日本語文の入出力機器の開発，または漢字処理システムの開発にあたって，最初に遭遇する難問の1つは，使用文字種類の決定である．従来使われている，いわゆる漢字テレタイプライタのけん盤，印刷機においても，文字の種類はユーザごとに相違しており，総字種約 1,000～2,500 の範囲で各種のものが用いられている．標準の字種というものは決まっていない．

われわれは漢字表示装置の文字種類を決めるための資料として，最近の日刊紙における文字使用状態の調査を行なったので，その概要と結果の一部を報告する．

2. 調査の計画

毎日新聞社では，漢字けん盤さん孔機を用いて，主要紙面の紙テープを作成し，これをモノタイプにかけて紙面をつくるとともに，地方本社に伝送して，全国紙面を各地方本社で作成している（この紙テープを幹線テープという）．そこで，この幹線テープを毎日新聞社より借用して，計算機で集計する方法を計画した．新聞紙面における文字使用状態には季節変動があるので，丁度1年間を調査対象期間とした．しかし，1年間の全紙面を対象にすると大変なので，1週間のうち，くせのない1日として木曜日を選び，その朝刊の全国紙面掲載分で紙テープのあるものを調査対象とした（このため，小説，家庭，娯楽，ラジオ，テレビが除かれた），期間は42年4月1日より43年3月31日までである．

集計は紙面のページ別に対応する7部門，①政治，②内政，③外電，④学芸・オピニオン，⑤経済・証券，⑥運動，⑦社会・対社，⑧総合，について行なった．

したがって，調査対象の文字種類は，毎日新聞社で用いている文字種類に限定された．漢字けん盤の標準字種（内字）は2,396で（**Table 1** 参照）．このほかに第1外字2,279種，第2外字がある．われわれは内字および第1外字を調査の対象とした．

Table 1　文字の種類

文字種類		字数		備考
仮名	ひらがな	77		濁音，半濁音，促音，拗音の文字も含む
	カタカナ	81	158	
漢字	当用漢字	1,818		当用漢字総数：1,850
	人名用漢字	63		
	地方向漢字	40		東京地方本社で使用されている漢字
	その他の漢字	171	2,038	
数字および符号	アラビア数字	100		0～99
	特殊数字	29		①～⑩，㈠～㈩，㊀～㊉
	アルファベット	26		A～Z
	符号	45	200	%，└，ルマ，トメ，イ，ル，ー その他
合計			2,396	

3. 実施方法

毎日新聞社より紙テープを運搬し，接着して長尺とした．これを計算機に入力して磁気テープに移し，処理・集計をした．以上を毎週1回行ない，1年間の集計の終了時に結果を印刷した．ここで用いた計算機はOKITAC-5090 C である．

4. 調査結果

ここでは内字についてのみ結果をしるす．調査結果を **Table 2** および **Fig. 1～3** に示す．

Table 2　調査字数

紙面	出現字数	割合(%)	文字の種類
① 政治	801,146	15.4	2,231
② 内政	639,076	12.4	2,193
③ 外電	711,843	13.7	2,178
④ 学芸，オピニオン	764,979	14.7	2,195
⑤ 経済，証券	513,966	9.9	2,244
⑥ 運動	678,179	13.0	2,209
⑦ 社会，対社	1,086,460	20.9	2,331
⑧ 総合	5,195,649	100.0	2,383

* On Frequency Distribution of Japanese Chracters in a Recent Newspaper by Hatsuho Hayashibara, Katsuhisa Yano, and Yoshiaki Kurosaki (Oki Electric Industry Co., Ltd.)
** 沖電気工業株式会社研究所

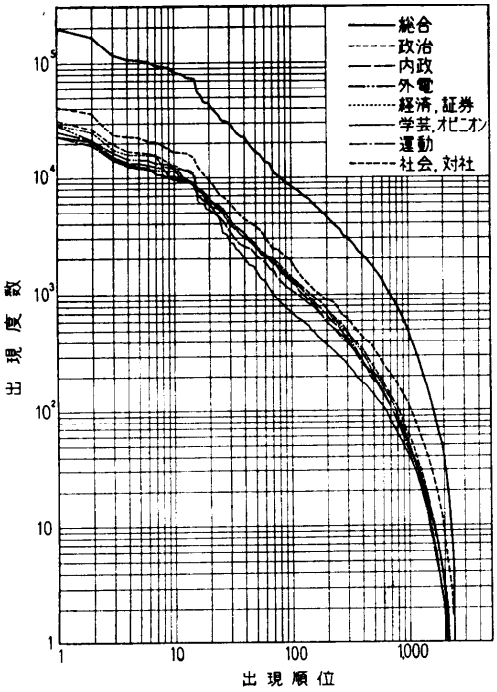

Fig. 1 出現順位-出現度数

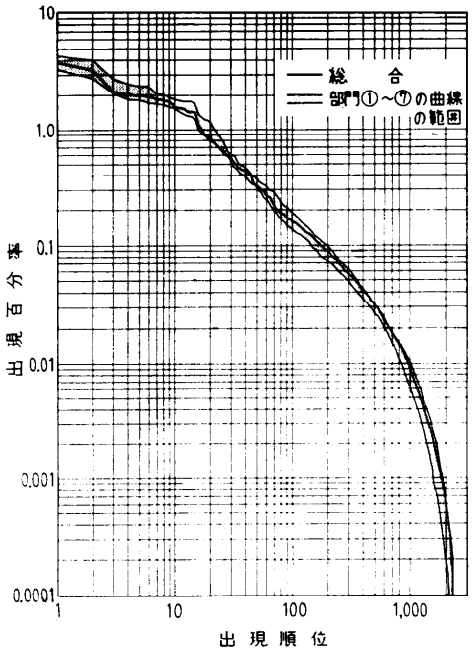

Fig. 2 出現順位-出現百分率

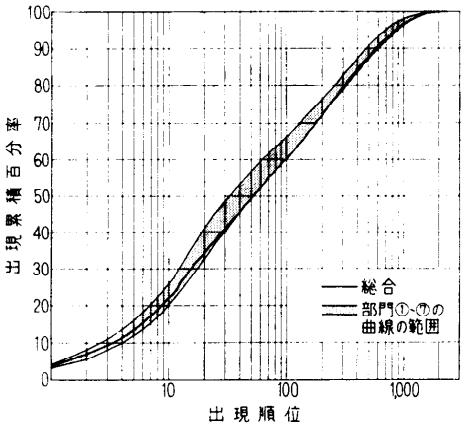

Fig. 3 出現順位-出現累積百分率

Fig. 3 より，総合では字種 600 で，累積百分率が 90%，1,500 で 99%，2,050 で 99.9% となることがわかった．また，部門①～⑦でやや性格の相違が認められた．たとえば，部門⑤学芸・オピニオンは，最も使用文字がかたよっていること，部門⑥運動は頻度の高い文字の出現率は比較的低いが，出現順位 100 位以降の立ち方が急激であることなどがわかった．

5. むすび

文字使用状態の調査は，従来各所で行なわれており，最も大がかりでまとまっているものには，毎日，朝日その他の新聞社の活字調査，および国立国語研究所の現代雑誌 90 種の用語用字調査などがある．しかし，これらは比較的古く，また，それぞれに調査方法が異なっているので，直接にわれわれの調査結果と比較することができない．

漢字けん盤に含まれている 2,396 字種のうち，この調査で一度も使われていなかった文字もあり，いわゆる，外字を含めても 2,500 種以下で，通常の使用状態では，ほとんど問題のないことがわかった．ただし，使用頻度の低い字で，一般に最も問題となるものは，地名と人名であるといわれており，これらを完全にカバーするためには，2,500 字では足りない．一般に，このような場合には 5,000～6,000 字が必要といわれているが，これらについては，現代日本の地名，人名に関する別途の調査が必要と思われる．

なお，本調査は通産省大型プロジェクト（超大型電子計算機の開発）の「陰極線管式漢字表示装置に関する研究」の一部として行なわれたものである．

謝 辞

本調査にあたっては，各方面の方々のご協力を得た．まず，計画にあたっては，当時国立国語研究所第四部長の林 大氏，および当時沖電気研究所嘱託をお願いしていた佐藤敬之輔氏から有益なご指導をいただいた．さらに，計画と結果について，毎日新聞社古川恒氏から懇切なご意見をいただいた．また，計画・実施・結果と全般にわたって，企画調査局を中心として毎日新聞社各位の全面的なご協力を得た．最後に，毎日新聞社との連絡，紙テープの運搬に，当社データ処理営業部の協力を得た．以上しるして謝意を表する．

参考文献

1) 日本新聞協会工務委員会：新聞印刷活版編，pp. 40～49，社団法人日本新聞協会，昭和40年．
2) 毎日新聞社東京工務局：本社使用活字頻度調査，昭和24年．
3) 岩淵悦太郎，他：現代雑誌九十種の用語用字，第1分冊，秀英出版，昭和37年．

（昭和44年4月8日受付）

雑 報

IFAC 京都シンポジウム

「IFAC Symposium on Systems Engineering Approach to Computer Control」が **1970年8月11日～14日**に国立京都国際会館で開催されます．内容は

（1） Methodology for system identification and optimization，（2） Uncertainty in systems and stochastic processes，（3） Adaptability of computer control，（4） Computer aided control in industrial aud other systems，（5） Computer aided system design methods，（6） Economics of computer controlled systems.

の6項目にわたっています．論文要旨の申込み受付けは，**1969年10月末日**までとなっています．論文申込あるいは参加希望者は，「京都市左京区吉田河原町 14 日本自動制御協会気付 IFAC 京都シンポジウム委員会」宛に返信用封筒を同封して「First Circular, Call for papers」を請求して下さい．

第6回 AICA Congress

AICA-IFIP 共催の Hybrid Computation に関する Conference が **1970年8月31日～9月4日**にミュンヘンで開催されます．論文要旨申込み期日は **1969年10月末日**までとなっていますので，詳細を知りたい方は学会へ問合わせてください．なお Scientific Programme は下記のとおりになっています．

（1） Fundamental methods for hybrid computation，（2） Applications，（3） Hybrid system operation，（4） Components of hybrid systems.

「各種産業における生産管理情報システム」に関するセミナー

日本自動制御協会（京都市左京区吉田河原町 14）では，来る10月21日～22日に日本都市センター（東京都千代田区平河町 2—6）で標記のセミナーを開催します．申込みは先着150名となっており，受講希望者は，同協会に至急お申し込み下さい．なお，受講料は本学会員も会員特典 4,500円（一般 5,500円）を受けられるので，その旨お申し出ください．