

漢字の性質について*

多 田 健 一**

近年情報処理の発達に伴い、わが国独自の国語である日本語の取扱い、中でも漢字処理の問題が、急速にクローズアップされてきた。この際、漢字をどのように取り扱うかという課題にとりくむにあたって、漢字の基本的性質を調べるのが、いずれにしても重要であると思われる。

そこで、手はじめに日常もっとも使用されている明朝体をとり上げ、これを画素に分解して検討を試みた。

1. 調査対象文字と分析手法

a) 調査対象文字

現代雑誌 90 種の用語用字(2)***の漢字使用率表に基づき検討した結果、使用漢字 3,328 種中、上位 2,000 字でほとんど 100% をカバーしていることがわかったと同時に、それら 2,000 字中に、当用漢字・人名用漢字・補正案追加漢字の大部分ははいっていることから、われわれの当面の分析対象文字を下表のようにきめた。

教育漢字	881
当用漢字(教育漢字をのぞく)	969
人名用漢字	92
補正案追加漢字	26
(28 字中人名用漢字と重複する字「杉、尚、齊」をのぞき、「燈」の略字体である「灯」を入れた。)	
合計	1,968 字

b) 分析方法

画素数は、分解しても文字の形がくずれない程度のメッシュとして $30 \times 30 = 900$ とした。

$30 \times 30 = 900$ の画素で表現された文字イメージをつくり、タテ方向の 1 列ごとに、連続した白(または黒)の画素を次の黒(または白)の画素にぶつかるとまで計数して 1 桁 (digit) とする。これを 30 列について順次実行して得られる digit を 1 文字イメージごとに記録し集計する。同じ手順をヨコ方向についても行なう。

図 1 では「高」という文字イメージのヨコ方向の例

をあらわしたものだが、最右端の数字はヨコ 1 列の桁数で、それらを全部加えると 156 digit となる。すなわち、この文字をヨコに分解したときの白と黒のときの数(ふちも一つのときれとする)が 156 であることをあらわす。これは文字の複雑さをあらわす数字としての意味をもつ。

2. 統計

このようにして分析した結果は表 1 のようになる。これからわかることは、全般的に、ヨコ方向に比してタテ方向に分析した方が桁数が多くなる傾向にある。すなわち、漢字はおおむねタテ方向の線素より、ヨコ

表 1

		タテ (digit)	ヨコ (digit)
教 育 (881 字)	全 体	171,724	154,663
	1 文字平均	194.9	175.6
当 用 (969 字)	全 体	197,152	187,184
	1 文字平均	203.5	193.2
人 名 用 (92 字)	全 体	18,854	17,432
	1 文字平均	204.9	190.3
補 正 (26 字)	全 体	5,066	5,056
	1 文字平均	194.8	194.3
合 計 (1,968 字)	全 体	392,796	364,335
	1 文字平均	199.6	185.1

方向の線素の方が多いということがいえる。これは図 2 からいえる。

図 2 は 20 桁ごとに区切って文字数の分布、すなわち、桁数と文字数との関係をみた指向特性曲線である。なお、人名用 92 字、補正追加 26 字については、文字数が少なく特性があらわれにくいではふいた。

図 2 から非常に興味あることがわかった。すなわち教育漢字グループと、教育漢字をのぞいた当用漢字グループを比較してみると、ヨコの指向特性曲線の形は異なるにもかかわらず、タテ方向のそれはほとんど相似であることがわかる。これは明朝体の特徴である“ウロコ”* が原因していると思われる。いいかえると、“ウロコ”は横方向分析に与える影響が非常に大き

* Characteristics of KANJI by Ken-ichi TADA (Research & Development of TAKACHIHO KOEKI Co., Ltd.)

** 高千穂交易株式会社研究開発部

*** 国立国語研究所出版

* 明朝体活字で横線の右端の三角の山型

く、タテ方向にほとんど影響を与えないからである。

教育漢字グループをみると、ヨコの指向性は、全体に左右対称となり、平均して 180 digit 近辺にその中心がある。タテは 190 にピークがあり、こころもち縦長の指向性を示す。

教育漢字をのぞいた当用漢字のグループでは、タテはやはり 190 にピークがあらわれ縦長であるが、ヨコは 210 にピークがあらわれ、全体に教育漢字グループに比して右寄り、すなわち、複雑さが増していること

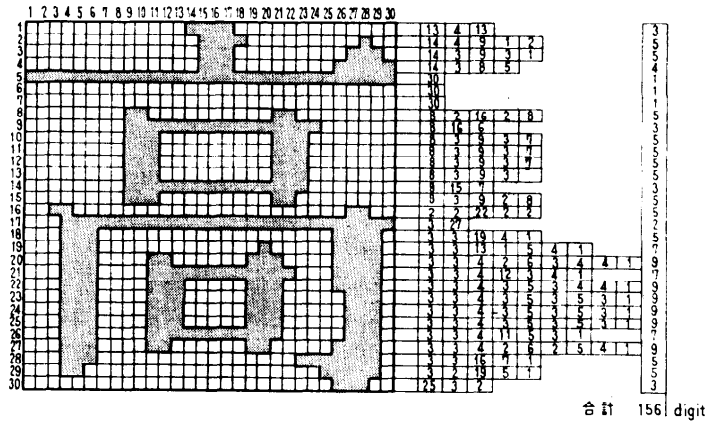


図 1 漢字のメッシュ分解

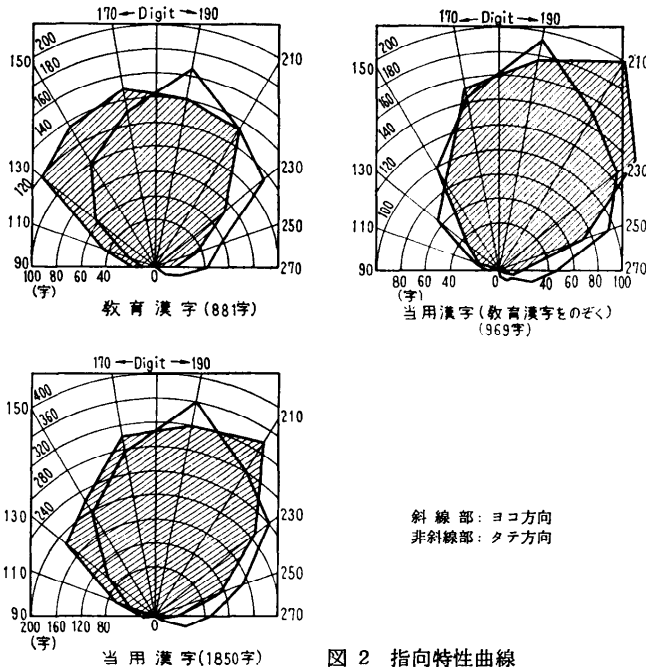


図 2 指向特性曲線



図 3

がうかがえる。

当用漢字全体では、タテ方向指向は部分グループ(教育漢字グループと教育漢字以外の当用漢字グループ)のそれらとまったく同形にあらわれ、ヨコ方向指向はピークが 210 にあるが、全体としてはタテに比して左寄りに傾いている。なお、これら指向特性グラフにおいては、グラフでかこまれた面積はあまり重要な意味はもっていない。

今回は明朝体について分析したが、もし“ウロコ”のないゴシック書体について分析したら、タテの指向曲

線はあまりこの図とかわからないが、ヨコの指向曲線は全体にもっと左にかたむき、タテ・ヨコの指向性が、さらにはっきり分けられると予想される。

3. 最後に

文字イメージを 30×30 のメッシュに分離して記憶し、再び Display してその文字品位をみたのが図 3 である。

ほとんど最初われわれが予想したとおり、明朝体活字を無理なく表現できることがわかる。