

国立国語研究所学術情報リポジトリ

The Total Vocabulary and Their Written Forms in Ninety Magazines of Today : Manual

メタデータ	言語: jpn 出版者: 公開日: 2019-11-29 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00002344

現代雑誌九十種の用語・用字全語彙・表記

国立国語研究所言語処理データ集 7

現代雑誌九十種の 用語用字 全語彙・表記

【FD版】

国立国語研究所

三省堂

-BK

31

904232

国立国語研究所言語処理データ集 7

現代雑誌九十種の用語用字 全語彙・表記

【FD版】

●——収録データ

国立国語研究所が1956年に雑誌九十種を対象として行った用語用字に関する調査結果。
レコード件数 4万。
各見出し語の語種・品詞・表記・度数。
総量 1MB。

●——商品構成

3.5インチ (1.44MB) フロッピーディスク 1枚
(データ形式 MS-DOSテキスト)
解説書 1冊

■発行 株式会社 三省堂

〒101 東京都千代田区三崎町2丁目22番14号

電話 編集 (03)3230-9411

販売 (03)3230-9412

© The National Language Research Institute 1997



T4938641613005

■定価 (本体4,000円+税)

ISBN4-385-61300-1 C3881 ¥4000E

国立国語研究所



1001904232

刊行のことは

国立国語研究所は、現代日本語に関する多様な研究を行ってきています。書きことばについては、用語・用字に関する研究が積み重ねられています。

その一つに、昭和31年当時の代表的な一般雑誌九十種に関する調査研究があります。その成果は、すでに

国立国語研究所報告21・22・25

『現代雑誌九十種の用語用字 第一・二・三分冊』（昭和37-39年）
として刊行し、また、用いたカードは、マイクロフィッシュによって、
国立国語研究所言語処理データ集3

『現代雑誌九十種の用語用字 五十音順語彙表・採集カード』（昭和62年）
として公表しました。語彙全部を一覧し、その表記や文脈を知ることは、このマイクロフィッシュによってできるようになりました。

このたび、語彙全部の一覧およびその表記を、度数つきで、機械可読形式で提供します。これによって、現代日本語に関する言語学的研究、あるいは数理的な研究、情報処理的研究、その他種々の研究が、一層発展することを、願ってやみません。

なお、このデータの整備には、この調査研究に携わった宮島達夫（現在、京都橘女子大学教授）が当たりました。

平成9年3月

国立国語研究所長 水谷 修

現代雑誌九十種の用語用字

全語彙・表記

データ解説

1 ——このデータの意義

このデータは、1956（昭和31）年の雑誌九十種の用語・用字について国立国語研究所が調査した結果の一部を、電子媒体にのせたものである。

この調査については、すでに、「国立国語研究所報告」として、つぎの報告が刊行されている。

報告21『現代雑誌九十種の用語用字 第一分冊 総記および語彙表』

（1962年、秀英出版発行）

報告22『現代雑誌九十種の用語用字 第二分冊 漢字表』

（1963年、秀英出版発行）

報告25『現代雑誌九十種の用語用字 第三分冊 分析』

（1964年、秀英出版発行）

また、「国立国語研究所言語処理データ集」として、つぎのマイクロフィッシュが刊行されている。

データ集3『現代雑誌九十種の用語用字 五十音順語彙表・採集カード』

（1987年、東京都板橋福祉工場発行）

さらに、「国立国語研究所研究報告集」には、この資料を分析した、つぎの報告がのっている。

宮島達夫・高木翠「雑誌九十種資料の漢語表記」

研究報告集 1（国立国語研究所報告 62, 1978年、秀英出版発行）

宮島達夫「意味分野と語種」

研究報告集 2（国立国語研究所報告 65, 1980年、秀英出版発行）

宮島達夫・高木翠「雑誌九十種資料の外来語表記」

研究報告集 5（国立国語研究所報告 79, 1984年、秀英出版発行）

宮島達夫・高木翠「雑誌九十種資料の和語表記」

研究報告集12（国立国語研究所報告103, 1991年、秀英出版発行）

ここに刊行するデータをあつかった報告として、つぎがある。

宮島達夫「雑誌九十種表記表の統計」

日本語科学 1 (1997年, 国書刊行会発行)

この調査対象は40年もまえのものであって、いまとなつては、〈現代〉とよぶには、ふさわしくない。それでもこのデータを刊行することには、つぎのような意義がある。

- 1) 報告21にのっている語彙表は、使用された度数が7以上のものであり、度数6以下のものをふくんで全語彙の度数をしるした語彙表は、印刷物の形では発表されていない。マイクロフィッシュによるデータ集3には、全語彙の度数表がはいっているが、利用しやすいとはいえない。
- 2) このデータでは、単語の使用度数だけでなく、その表記ごとの度数をしるした。語形のゆれもとりあげた。このような形での表記表は、発表されていない。
- 3) すべての見出し語に語種と〈品詞〉とをしるしてあり、4万語の見出し語をもった一種のコンピューター用辞書として利用することが可能である。(ここでいう〈品詞〉については、あとの説明を参照。)

報告21には、助詞・助動詞の度数表もあるが、これはこのデータでは省略した。また、報告21の語彙表には○●などの符号の度数もふくまれているが、これも今回ははぶいた。

本調査の対象である雑誌は、つぎの五つの〈層〉に分類されていた。

- | | |
|------------|---------------------------|
| 1層 (評論・芸文) | 世界, 中央公論, 群像... |
| 2層 (庶民) | 家の光, 週刊朝日, リーダーズダイジェスト... |
| 3層 (実用・科学) | エコノミスト, 科学朝日, 東洋経済新報... |
| 4層 (生活・婦人) | 主婦の友, それいゆ, 暮らしの手帖... |
| 5層 (娯楽・趣味) | 映画の友, オール読物, 相撲, 明星... |

報告21の語彙表では、これらの各層ごとに使用率がしめされている。しかし、このデータ(およびデータ集3収載の語彙表)では、層別の内わけを無視して、各層の合計度数だけをしめた。

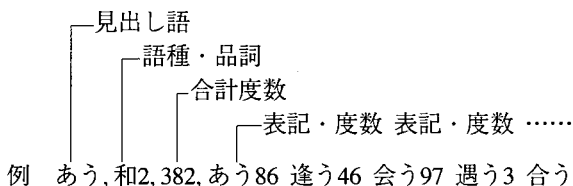
2——データの形式

このデータは、zassi90.txt となづけられた MS-DOS テキストファイルである。レコード件数39,997からなり、総量1,022,179バイトである。

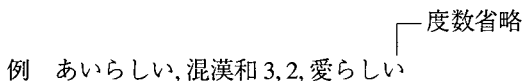
このデータの各レコードは、つぎの4フィールドからなる。記述にもちいた文字の種類も、あわせてしるす。

- 1) 見出し語 全角 (2バイト) カタカナ
- 2) 語種・品詞 語種は全角 (2バイト) 文字,
品詞は半角 (1バイト) 数字
- 3) 合計度数 半角 (1バイト) 数字
- 4) 表記とその度数 表記は全角 (2バイト) 文字,
度数は半角 (1バイト) 数字

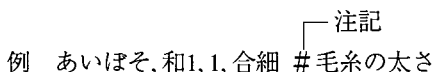
各フィールドは半角のコンマ「,」でくぎった。語種と品詞はつづけ、表記とその度数もつづけた。表記の種類が二つ以上あるときには、表記・度数のくみをくりかえし、そのあいだに半角の空白「 」をおいた。



表記の種類が一つであるときには、表記別の度数を省略した。



第4フィールドに、注記をくわえたものがある。表記とその度数のあとに、半角空白「 」と全角の「#」をおいて、しるした。



以下、フィールドの順序と逆になるが、表記、語種・品詞、見出し語について、注意すべきことをしるす。

かなと漢字とのちがいが、カタカナとひらがなのちがいが、おくりがなのちがいが、漢字どうしでのちがいは、表記の種類が別であるとした。しかし、歴史的かなづかいは現代かなづかいに合併し、異体の漢字は通用の字体に合併した。

例 あおる、和2,7,あおる3 アオル1 煽る3

あかるい、和3,103,あかるい2 明かるい5 明るい96

「むずかしい～むつかしい」のような語形のゆれも、表記が別であるとして、とりあげた。しかし、活用形のちがいはとりあげていない。

原文にあるルビのうち、必要な（それがないと、ほかの読みの可能性がある）ばあいには、く > でかこんでしめた。ただし、ルビがなくても、ほかに読まれるおそれがなときは、原文にルビがあったことをしめていない。

例 あいのこ、和1,2,アイノ子1 混血児くあいのこ> 1

このデータでの〈語種〉とその略称は、つぎのとおりである。

和語 → 和
漢語 → 漢
外来語 → 外
混種語 → 混
人名 → 人／姓／名／姓名
地名 → 地

人名のうち、「人」は、略称の「Q」のように、人名であることはたしかであるが、姓か名か不明のもの、「姓名」は「エノケン」のように、姓と名とを結合したものである。

語種が問題になるものの例をあげると、「尼・あばた・瓦・さじ・みそ」などは和語、「絵（え）」も和語としたが、「図絵」は漢語とした。「カルマ・インモ・ゆず」は漢語とした。

外来語には原則として原語をあげた。しかし、会社名・製品名など、いくつかについては、形からみて外来語と考えられても原語は見当がつかないことがあり、それらのばあいには、単に「外」としてした。また、原語のうち、以下のものは略称をつかった。

英語 → 英 ドイツ語 → ド フランス語 → フ
イタリア語 → イ ロシア語 → ロ オランダ語 → オ
ポルトガル語 → ポ スペイン語 → ス ラテン語 → ラ
アイヌ語 → ア 中国語 → 中 朝鮮語 → 朝

このデータで〈品詞〉とよぶのは、国立国語研究所資料集6『分類語彙表』（1964年、秀英出版発行）の4分類で、それぞれ数字で表現している。ただし、人名・地名は、ほんらい体の類であるが、1という数字を省略した。

- 1 体の類（名詞）
- 2 用の類（動詞）
- 3 相の類（形容詞・形容動詞・副詞）
- 4 その他の類（接続詞・感動詞・陳述副詞）

文章のなかでは、おなじ単語が名詞としても副詞としてもつかわれる、ということがあるが、このデータでは、便宜上、1語には一つの品詞名しかあたえなかった。しかも、じっさいの用例をみて品詞を決定したわけではないので、実例は名詞的な用法だけなのに、3という分類になっている、というような可能性もある。

見出し語の配列は、いわゆる国語辞典順の一種である。大字の「つ」「や」などは小字の「っ」「ゃ」などよりもまえにおいた。この配列は、豊島正之氏が作成したフリーソフトウェアsortf（旧版）が採用しているものであり、じっさい、この配列にはsortfをつかった。

例 規約（きやく） → 客（きゃく）

同音語の配列は、品詞によって

体 → 人名 → 地名 → 用 → 相 → その他

とし、おなじ品詞では、語種によって

和 → 漢 → 外 → 混

とした。

見出し語の立て方は、報告21では独特の方法をとっているばあいがあるが、ここでは一般に採用されている方式によることとして、カタカナ表記をひらがな表記とし、また、つぎの3点を変更した。配列も報告21とちがう点が生じた。

- 1) 動詞の代表形を、報告21では連用形にして、例えば「見る」「来る」を「ミ」「キ」の形とし、それにしただった位置に配列したのであるが、ここでは終止形にして、それらの動詞を「みる」「くる」の位置においた。
- 2) 形容詞の代表形を、報告21では語幹としたが、ここでは終止形にした。これは、語尾が「い」であるために、ほとんど実質的な変更にはならないが、たとえば、報告21では

赤 → 赤い → あかあか（と）

の順にしたところが、ここでは

赤 → あかあか(と) → 赤い

の順にした。

- 3) 「学校」「石器」などの語形を、報告21では「ガクコウ」「セキキ」などと表記したが、ここでは「がっこう」「せっき」とした。配列もことなる。

3——単位について

調査単位との関係で、注意すべき点をしるす。調査単位のくわしい規定は、報告21および25を参照していただきたい。

- 1) 常識的な意味での単語がデータになく、もっとこまかい単位にわかれていることがある。たとえば、「自動車」「飛行機」「小学校」などの度数があがっていないが、これらの用例がなかったのではなく、それぞれ「自動／車」「飛行／機」「小学／校」と分割して、単独の「自動」などに合併して集計したのである。
- 2) 「入出国」の類は、「入国」1、「出国」1としてかぞえてある。
- 3) 接辞も単語なみにあつかわれている。それで、「おとうさん」という単語はなく、「お／とう／さん」という3単位に分割されている。
- 4) 「あそび／くらす」「あばれ／まわる」などは複合動詞として1単位になっている。しかし、生産力のたかい接尾的動詞・形容詞は接辞あつかいであり、「あそび／はじめる」「あばれ／だす」や「あつかい／にくい」などは、それぞれ、きれている。同様に、「読み／方」も2単位であり、このさい、「読み」は動詞としてある。助数詞も、「五／人」「八／杯」などで、数とは別の単位になる。
なお、これら後要素が連濁をおこしていることがあるが、それらは連濁のない形と合併してある。たとえば、助数詞「杯」には「ばい」「パイ」などもあるが、これらは「はい」「ハイ」とあわせてかぞえてある。
- 5) いわゆる助動詞のうち、「(ら)れる」「(さ)せる」「(たい)は接尾辞と同様に、動詞・形容詞としてあつかっている。
- 6) 可能動詞（「読める」など）は、もとの動詞（「読む」など）に合併してある。
- 7) 接尾辞「さ」のついた形は、派生名詞としないで形容詞に合併してある。

以上のような規定は、このデータをつかうばあいには、している必要がある。たとえば、動詞「読む」の度数には、ふつうの意味での動詞のほかに、「読み方」や「読みはじめる」での度数も合算されている。動詞「ぬう」の表記には、「ぬう」や「縫う」のほかに「縫」もあるが、おくりがなのないこの形は、漢文の白文にでてきた可能性よりも、「縫い方」の前要素である可能性がたかく、それら異質のものを合併して、「ぬう」のおくりがなを論ずることは危険である。(なお、動詞でおくりがなのない形には、「不可侮」のような、漢文の白文にでてくるものもある。)

同語か異語かの判別については、このデータだけではわかりにくいものもあり、報告25の294ページ以下に用例をあげて説明してあるので、参照されたい。

国立国語研究所言語処理データ集 7
現代雑誌九十種の用語用字 全語彙・表記
[FD版]

The Total Vocabulary and their Written Forms in
NINETY MAGAZINES OF TODAY

平成9年9月1日 第1刷発行
著作者 国立国語研究所
発行者 株式会社三省堂

発行所 株式会社三省堂
〒101 東京都千代田区三崎町2丁目22番14号
電話 編集 (03) 3230-9411
販売 (03) 3230-9412

Av
B
1001

F

110611 000000