

[フロッピー版] 中学校・高校教科書の語彙調査 国立国語研究所編 解説書

著者	国立国語研究所
ページ	5-19
発行年	1994-06
シリーズ	国立国語研究所言語処理データ集 ; 6
URL	http://doi.org/10.15084/00002343

国立国語研究所言語処理データ集 6

[フロッピー版]

中学校・高校教科書の語彙調査

国立国語研究所 編

解説書

秀英出版

-BK

30

13538



国立国語研究所



1002013538

刊行のことば

国立国語研究所言語処理データ集6として、『中学校・高校教科書の語彙調査』〔フロッピー版〕を刊行します。

先に刊行した『高校教科書の語彙調査』2冊、『中学校教科書の語彙調査』2冊は、語彙表を主体としたものでした。本データ集はこれをほぼそのままフロッピーにうつし、コンピュータによって利用できるようにしたものです。

この調査は、国民が一般教養として、各分野の専門知識を身につける時に必要と思われる語彙の実態を明らかにすることを目的として企画しました。その調査結果である先の報告書は、専門語研究、理科・社会科教育、日本語教育、辞書作成など各分野で利用されました。

最近では、多くの研究者がパソコンなどを用いてさらに深い分析を行い、またワープロや機械翻訳などの辞書にも専門語を入れ、専門分野の高度な処理を行うようになりました。現在、いろいろな分野で頻度付き単語データの電子媒体での公開を求める声が聞かれます。このデータ集がそれぞれの研究分野の多くの方々に利用されることを望みます。

本データ集の作成は、言語体系研究部第2研究室の協力のもとに主として中野洋（前言語体系研究部第2研究室長、現言語体系研究部長）が担当し、山崎誠（言語体系研究部主任研究官）、小沼悦（同第2研究室研究補助員）がこれを助けました。

平成6年3月1日

国立国語研究所長 水谷 修

『中学校・高校教科書の語彙調査』[フロッピー版]

解 説 書

1. データ公開の趣旨

国立国語研究所では、これまで雑誌、新聞、教科書等を対象として大量語彙調査を行ってきた。ここに公表するのは高校教科書、中学校教科書の理科、社会科の語彙調査結果である語彙表をフロッピーにうつしたものである。調査の目的、調査対象、方法、語彙量などについてはこのフロッピーにも収めたそれぞれの報告書の「調査の概要」を参照されたい。

語彙表は、すでに下記に示す4冊の報告書として刊行した。

『国立国語研究所報告76 高校教科書の語彙調査』 (秀英出版, 1983)

『国立国語研究所報告81 高校教科書の語彙調査Ⅱ』 (秀英出版, 1984)

『国立国語研究所報告87 中学校教科書の語彙調査』 (秀英出版, 1986)

『国立国語研究所報告91 中学校教科書の語彙調査Ⅱ』 (秀英出版, 1987)

またその分析結果を次の報告書に公表した。

『国立国語研究所報告99 高校・中学校教科書の語彙調査 分析編』

(秀英出版, 1989)

以上の報告書は、専門語研究、高校や中学校の理科・社会科教育の現場、日本語教育の教材作成、国語辞典やワープロ、機械翻訳などの辞書作成など各分野で利用された。

ところで、語彙表は度数順と五十音順に配列した語のリストである。ある語の使用率を求めるには五十音順語彙表を調べればよい。同様に教科書の上位100語を求めるには度数順語彙表を調べればよい。しかし、たとえば外来語の使用率やある意味分野の語の使用率を求めるためには語彙表のすべてを調べなければならない。またある語構成要素を含む語を求めるには、五十音順や度数順語彙表では困難である。さらには、中学校教科書と高校教科書の語彙の比較も先の報告書を用いての作業は大変である。このために、『電子計算機による新聞の語彙調査Ⅱ』では語種別、品詞別語彙表を、『電子計算機による新聞の語彙調査Ⅲ』では語構成を示す接続表を報告書として刊行した。

これらは、コンピュータを用いれば比較的簡単に作成することができる。当時まだ国語研究にコンピュータを使うことはまれであったので、それぞれの報告書にその方法を詳しく記した。

最近では、文科系の研究においてもコンピュータを用いることが多くなった。また、言語情報処理の研究では大規模な言語データが必要となり、機械処理用辞書の作成が各所で行われている。これを反映して、国立国語研究所の『分類語彙表』や各種の頻度付き語彙表の公開を求める声が大きくなったのである。

本データ集は、これらの声にこたえ、先の『分類語彙表』[フロッピー版]に続いて、『中学校・高校教科書の語彙調査』[フロッピー版]として刊行するものである。

本データは、報告書と異なる。大きな点では、本データは五十音順語彙表だけであるが、語彙表では度数順語彙表なども掲載されている。五十音順語彙表と度数順語彙表は、データの並べ方が異なるだけであって、コンピュータ処理すれば自動的に作成することができるためである。また、細かい点では、表の罫線がない、表の桁あわせがされておらず区切り記号のカンマ(,)で示している、項目の順序が異なるなどである。これらについては第6章のデータフォーマットを参照されたい。さらに本データを機械処理する人のために代表形や判別情報などの報告書にはない情報を付けた。このことについては報告書の「調査の概要」を参照されたい。

しかし、いずれもおおもとの内容は同じである。

2. 高校教科書の語彙調査と中学校教科書の語彙調査の概要

2つの語彙調査の目的、調査対象、調査単位について、各報告書の概要から引用して概略説明する。詳しくは、報告書または概要のファイルを参照いただきたい。

2.1 目的

2.1.1 高校教科書の語彙調査

この調査は、国民の一般教養として、各分野の専門知識を身につける時に

必要と思われる語彙の実態を明らかにすることを目的として企画された。高等学校進学率の増加に伴い、現今では、高等学校教育は、国民大多数の基本的な教養の場となっている。また、大学教育は、この高校教育の基盤に立って進められるものであり、とくに高校の理科と社会科は、大学における専門教育の基礎となっていると考えることができる。われわれが高校の理科・社会科の全教科を対象とした語彙調査を企画したのは、以上のような理由からである。

2.1.2 中学校教科書の語彙調査

この調査は、国民の一般教養として、各分野の専門知識を身につける時に必要と思われる語彙の実態を明らかにすることを目的として行われた。

中学校教科書調査は、高校教科書調査のあとを受けて、それより一段階前の、義務教育の最終段階である中学校で身につける知識体系を記述する語彙の実態を明らかにすることを目的としている。

2.2 調査の対象

高校教科書の語彙調査

(教科名)	(教科書名)	(著者名)	(出版社名)	(発行年月日)
物理 I	標準高等物理 I	大塚明郎ほか	講談社	昭和49年1月30日
化学 I	化学 I	柴田雄次ほか	大日本図書	昭和49年2月5日
生物 I	生物 I	石田寿老ほか	清水書院	昭和50年2月15日三版
地学 I	地学 I	湊 正雄ほか	実教出版	昭和49年1月25日
倫理社会	倫理・社会	中村 元ほか	東京書籍	昭和49年2月10日
政治経済	政治経済 新訂版	辻 清明ほか	自由書房	昭和49年2月5日
日本史	詳説日本史(再訂版)	宝月圭吾ほか	山川出版	昭和49年3月5日
世界史	三省堂新世界史	土井正興ほか	三省堂	昭和49年3月30日再版
地理 B	高校新地理 B	青野壽郎ほか	二宮書店	昭和49年1月20日

中学校教科書の語彙調査

(教科名)	(教科書名)	(著者名)	(出版社名)	(発行年月日)
理 1 上	新理科 1 分野上	坪井忠二ほか	大日本図書	昭和55年2月5日
理 1 下	” 1 分野下	”	”	”
理 2 上	” 2 分野上	”	”	”
理 2 下	” 2 分野下	”	”	”
歴史	新しい社会[歴史]	鶴飼信成ほか	東京書籍	”
地理	” [地理]	”	”	”
公民	” [公民]	”	”	”

上記の教科書の本文部分を全て調査対象とした。本文部分には、表紙・目次・奥付・各ページのノンブル等は、当然含まれないが、さらに、次のものも、本文とはしなかった。

○巻末の索引・年表等

○図表・写真およびその周辺部分にある説明の言葉

○脚注

さらに、本文に付随する次のものも、本文部分から除いた。

○人名・地名の上下に付いているアルファベット表記・生没年

○下付きルビ(外国語や外国音、また別の読みなどを示すのに使われる)

2.3 調査単位

この調査では、社会科・理科の教科書という文章の性格を考慮に入れ、新たに決めた長短2種類の単位を用いた。長い単位は、文の構成にあずかる要素(いわゆる文節)にもとづく単位で、wordの頭文字をとってW単位と名付けた。短い単位は、語の構成にあずかる要素(いわゆる最小単位)にもとづく単位で、morphemeの頭文字をとってM単位と名付けた。また、単位切りの段階では、句読点その他の記号類も、一般の言葉と同等に扱っているが、それは単位切り作業と、それに続く一連の機械処理を的確かつ迅速に行うためであり、最終出力段階では、本表から除かれて、別途に集計されている。

3. 公開ファイルの概要

データは、フロッピー2枚に13のファイルを収めた。次の通りである。

1 枚目

KAISetu.DOC	この解説書
GOIRYO.DAT	各語彙表ファイルの語数
GAIYO.LZH	各報告書の「調査の概要」
KMG.LZH	「高校教科書 M単位五十音順語彙表」
KWG1.LZH	「高校教科書 W単位五十音順語彙表」
KWG2.LZH	「高校教科書 W単位五十音順語彙表」
KWG3.LZH	「高校教科書 W単位五十音順語彙表」
LHA.EXE	高圧縮書庫管理プログラム

LHA. DOC	高圧縮書庫管理プログラム解説書
2 枚目	
TMG. LZH	「中学校教科書 M単位五十音順語彙表」
TWG. LZH	「中学校教科書 W単位五十音順語彙表」
LHA. EXE	高圧縮書庫管理プログラム
LHA. DOC	高圧縮書庫管理プログラム解説書

このうち拡張子が LZH である 7つのファイルは吉崎栄泰氏作成のプログラム LHA を使用して作った圧縮ファイルである。それらは次の22のファイルを含む。

GAIYO. LZH	調査の概要などの圧縮ファイル
GAIYO_K1. TXT	「高校教科書の語彙調査」 調査の概要など
GAIYO_K2. TXT	「高校教科書の語彙調査Ⅱ」 調査の概要など
GAIYO_T1. TXT	「中学校教科書の語彙調査」 調査の概要など
GAIYO_T2. TXT	「中学校教科書の語彙調査Ⅱ」 調査の概要など
KMG. LZH	「高校教科書 M単位五十音順語彙表」の圧縮ファイル
KMG	高校教科書 M単位五十音順語彙表 自立語
KMJ	高校教科書 M単位五十音順語彙表 助辞
KMN	高校教科書 M単位五十音順語彙表 数字
KMS	高校教科書 M単位五十音順語彙表 記号
KWG1. LZH	「高校教科書 W単位五十音順語彙表」の圧縮ファイル
KWG1	高校教科書 W単位五十音順語彙表 自立語 「あ」～
KWG2. LZH	「高校教科書 W単位五十音順語彙表」の圧縮ファイル
KWG2	高校教科書 W単位五十音順語彙表 自立語 「しゃるる」～
KWG3. LZH	「高校教科書 W単位五十音順語彙表」の圧縮ファイル

KWG3	高校教科書	W単位五十音順語彙表 自立語 「ひっしゅうがっか」～
KWJ	高校教科書	W単位五十音順語彙表 助辞
KWN	高校教科書	W単位五十音順語彙表 数字
KWS	高校教科書	W単位五十音順語彙表 記号
TMG.LZH	「中学校教科書 M単位五十音順語彙表」の圧縮ファイル	
TMG	中学校教科書	M単位五十音順語彙表 自立語
TMJ	中学校教科書	M単位五十音順語彙表 助辞
TMN	中学校教科書	M単位五十音順語彙表 数字
TMS	中学校教科書	M単位五十音順語彙表 記号
TWG.LZH	「中学校教科書 W単位五十音順語彙表」の圧縮ファイル	
TWG	中学校教科書	W単位五十音順語彙表 自立語
TWJ	中学校教科書	W単位五十音順語彙表 助辞
TWN	中学校教科書	W単位五十音順語彙表 数字
TWS	中学校教科書	W単位五十音順語彙表 記号

圧縮ファイルをもとに戻す方法は第8章に記す。もとのファイルは、それぞれMS-DOSのテキストファイルである。次章に、各ファイルの内容やデータのフォーマットを示す。

4. 調査の概要などのファイル

各報告書の「刊行のことば」「調査の概要」「語彙量」である。以下にその目次によって内容を示す。

『高校教科書の語彙調査』

刊行のことば

I 調査の概要

1. 調査の目的
2. 調査の規模

2.1 調査対象

『高校教科書の語彙調査Ⅱ』

刊行のことば

I 調査の概要

1. 調査の目的
2. 調査の規模

2.1 調査の対象

- 2.2 調査の担当者
- 3. 調査の手順
- 4. 調査単位について
 - 4.1 調査単位の種類と長さ
 - 4.2 W単位の切り方 (概略)
 - 4.3 M単位の切り方
 - 4.4 単位切りの例
- 5. 同語異語判別
 - 5.1 同語異語判別作業の概略
 - 5.2 代表形のつけ方
 - 5.3 見出し語の示し方
 - 5.4 意味番号・語種情報
- 6. 機械処理システム
 - 6.1 機械処理システムの基本的態度
 - 6.2 国立国語研究所の機械環境
 - 6.3 機械処理システム
 - 6.4 エラーデータとその処理
 - 6.5 今後の語彙調査機械処理システムのために

II 語彙量

『中学校教科書の語彙調査』

刊行のことば

I 調査の概要

- 1. 調査の目的
- 2. 調査の規模
 - 2.1 調査対象
 - 2.2 調査の担当者
- 3. 調査の手順
- 4. 調査単位について
 - 4.1 調査単位の種類と長さ
 - 4.2 W単位の切り方 (概略)
 - 〈付〉助辞のリスト
 - 4.3 M単位の切り方
 - 4.4 単位切りの例
- 5. 同語異語判別

- 2.2 調査の担当者
 - 3. 調査の手順
 - 4. 調査単位について
 - 4.1 調査単位の種類と長さ
 - 4.2 W単位の切り方
 - 4.3 M単位の切り方 (概略)
 - 4.4 単位切りの例
 - 5. W単位の性格
 - 5.1 同語異語判別作業
 - 5.2 見出しの示し方
 - 6. 機械処理システム
 - 6.1 W単位調査の概要
 - 6.2 エラーデータの修正
 - 6.3 代表形の作成
 - 6.4 W単位語彙表作成
- 参考文献

II 語彙量

『中学校教科書の語彙調査II』

刊行のことば

I 調査の概要

- 1. 調査の目的
- 2. 調査の規模
 - 2.1 調査の対象
 - 2.2 調査の担当者
- 3. 調査の手順
- 4. 調査単位について
 - 4.1 調査単位の種類と長さ
 - 4.2 W単位の基本的性質
 - 4.3 W単位の切り方
 - 4.4 同語異語判別
 - 4.5 見出しの示し方
- 5. 機械処理システム

- 5.1 同語異語判別作業の概略
 - 5.2 代表形のつけ方
 - 5.3 見出し語の示し方
 - 5.4 意味番号・語種情報
 - 5.1 W単位調査の概要
 - 5.2 エラーデータの修正
 - 5.3 ファイル形式
6. 機械処理システム
- 6.1 国立国語研究所の機械環境
 - 6.2 機械処理システム
 - 6.3 今後の語彙調査機械処理システムのために

II 語彙量

II 語彙量

なお、以下の点で報告書と異なる。

- ①表や図、フローチャートなどは省略したものがある。
- ②アンダーラインは付けていない。
- ③報告書の中の点線の斜線は、「|」で代用した。
- ④数式は、次のように示した。

ルート記号は、 $\sqrt{\quad}$ の直後に()で括ってそれがかかる部分を示した。

上付き数字、下付き数字は1バイト数字で示した。

分数記号は1バイトの/で示し、また分母、分子が式の時()で括って示した。例を示す。

$\sqrt[3]{3}$ / $\sqrt{2(x^n) = a/b}$ / $2H_2 + O_2 \rightarrow 2H_2O$ /

$a = (x + y) / 2$ /

%。(パーミル(千分率))

- ⑤「鶴岡昭夫」とした「鶴」は、正しくは「鷺」である。

いずれも、正しくはそれぞれの報告書を参照されたい。

5. 語彙表のデータフォーマット

5.1 高校教科書 M単位五十音順語彙表(自立語 助辞 数字 記号)

4つのファイルKMG, KMJ, KMN, KMS は、同じフォーマットである。

「高校教科書 M単位五十音順語彙表」のファイルは、1レコードが次の19の項目からなり、コンマ(,)で区切っている。項目を()でくくったものはデータにはない。語種は、和語を和、漢語を漢、外来語を外、混種語を混、人名を人、地名を地、助辞をJ、数字をN、記号をSで示す。(以下同じ)

項目はすべて可変長である。

KMGのフォーマットを示す。

表中のデータ例のレコードは次の通りである。

あかるい, 明, 3501, 明るい, 和, 39, .121, 1362.5, 6, 1, 7, 16, 30, .27, 4, 4, 1, ., 9, .04

語彙表例は、報告76の「五十音順M単位語彙表」の「あかるい」ではどう書いてあるかを示す。

項目（括弧内は語彙表）	データ例	語彙表例	備考
代表形	あかるい	なし	
（見出し）	なし	あかるい	語種情報を使えば語彙表の見出しが作成できる。
判別情報	明	なし	同じ代表形をこれで判別する。
意味番号	3501	3 5 0 1	
表記例・注記	明るい	明るい	
語種	和	なし	語彙表では和語と漢語は示さない。見出しの表記でわかる。
全体度数	39	39	
全体比率	.121	.121	桁上がりは、データでは R, 語彙表では ` で示す。千分率。
全体順位	1362.5	1362.5	
物理の度数	6	6	
化学の度数	1	1	
生物の度数	7	7	
地学の度数	16	16	
理科の度数	30	30	
理科の比率	.27	.27	
倫理社会の度数	4	4	
政治経済の度数	4	4	
日本史の度数	1	1	
世界史の度数			
地理の度数			
社会科の度数	9	9	
社会科の比率	.04	.04	

データには語彙表の見出しがない。しかし、代表形と語種情報を使えば機械的に生成できる。すなわち、見出しは、和語は代表形をそのままひらがな

で示し、漢語・外来語・混種語・人名・地名など代表形がひらがなならば片仮名で、数字や記号ならばそのままです。

語彙表には代表形と判別情報がない。これらは、機械処理で同語異語を明らかにするために用いたのであって、人間には見出しと表記例、語種によってどの語であるかがわかるからである。

5.2 高校教科書 W単位五十音順語彙表（自立語 助辞 数字 記号）

6つのファイル KWG1, KWG2, KWG3, KWJ, KWN, KWS は同じフォーマットである。

「W単位 五十音順語彙表」（自立語）は、KWG1, KWG2, KWG3 の3つのファイルに分けた。それらを1枚のフロッピーディスクにおさめるため1つのファイルの大きさを1メガバイト以内におさえたからである。KWG1, KWG2はそれぞれ15,000レコードである。

レコードは次の21の項目からなる。

KWG1のデータ例を示す。

ああ,あ,〔間投〕,.,3,.013R,8445.0,.....,3,....,3,.02R

あいする,愛,愛する,4,22,.094,1427.0,.....,19,2,1,.,22,.15R

あじあたいりく,ア,アジア大陸,7,.030R,4556.5,.,.,2,2,2.02,.,.,3,.,2,5,.03

語彙表の例は示さない。備考の通り、代表形と判別情報は語彙表にはない。データでは代表形と見出し表記例が同じ場合にこれを省略している。比率の中の最後の桁の記号Rは、最後の桁が四捨五入で繰り上がったことを示す。

項目	データ例1	データ例2	データ例3	備考
代表形	ああ	あいする	あじあたいりく	語彙表にはない。
判別情報	あ	愛	ア	語彙表にはない。
見出し表記例〔注記〕	〔間投〕	愛する	アジア大陸	代表形と同じ場合は省略。
語種				
表記数		4		
全体の度数	3	22	7	
全体の比率	.013R	.094	.030R	桁上がりは R。
全体の順位	8445.0	1427.0	4556.5	
物理の度数				
化学の度数				
生物の度数				

地学の度数			2
理科の度数			2
理科の比率			2.02
倫理社会の度数	3	19	
政治経済の度数		2	
日本史の度数		1	3
世界史の度数			
地理Bの度数			2
社会科の度数	3	22	5
社会科の比率	.02R	.15R	.03

5.3 「中学校教科書 M単位五十音順語彙表」(自立語 助辞 数字 記号)

4つのファイルTMG, TMJ, TMN, TMS は、同じフォーマットである。

「中学校教科書 M単位五十音順語彙表」のファイルは、レコードが次の19の項目からなる。

TMGのデータ例を示す。

あかるい, 明, 3501, 明るい, 和, 34, .258, 704, , 9, 19, 5, 33, .80R, 1, , , 1, .01

語彙表例は、報告87の「五十音順M単位語彙表」の「あかるい」ではどう書いてあるかを示す。

項目	データ例	語彙表例	備考
代表形 (見出し)	あかるい	あかるい	語彙表にはない。 データにはない。
判別情報	明		語彙表にはない。
意味番号	3501	3 5 0 1	
表記例・注記	明るい	明るい	
語種	和		語彙表では和語と漢語は示さない。
全体度数	34	34	
全体比率	.258	.258	
全体順位	703.5	703.5	
理科1上			
理科1下	9	9	
理科2上	19	19	
理科2下	5	5	
理科全体度数	33	33	
理科全体比率	.80R	.80	データでは桁上がりを入、 語彙表ではで示す。

公民の度数	1	1
地理の度数		
歴史の度数		
社会科の度数	1	1
社会科の比率	.01	.01

5.4 「中学校教科書 W単位五十音順語彙表」(自立語 助辞 数字 記号)

4つのファイルTWG, TWJ, TWN, TWS は、同じフォーマットである。

「中学校教科書 W単位五十音順語彙表」のファイルは、レコードが次の19の項目からなる。

TWGのデータ例を示す。

あいこくしん, 愛心, 愛国心, , 1, .010R, 12938.5, , , , , , , , 1, 1, .01
 けれども, 助, けれども, J, .2, .030, 41.0, 1, , , 1, 2, .09R, , , ,

TW.G01のレコードに含まれる項目は、次の通りである。

	データ例1	データ例2	備考
代表形	あいこくしん	けれども	語彙表にはない。
判別情報	愛心	助	語彙表にはない。
見出し表記例	愛国心	けれども	
表記数			
語種		J	
全体の度数	1	2	
全体の比率	.010R	.030	語彙表では桁上がりを で示す。
全体の順位	12938.5	41.0	
理科1上の度数		1	
理科1下の度数			
理科2上の度数			
理科2下の度数		1	
理科の度数		2	
理科の比率		.09R	語彙表では桁上がりを で示す。
公民の度数			
地理の度数			
歴史の度数	1		

社会科の度数 1
社会科の比率 .01

6. 二つの調査および報告書と本データとの異同

6.1 文字コードについて

「高校教科書の語彙調査」と「中学校教科書の語彙調査」の報告書は、ともに大型電子計算機の出力をそのまま印刷したものである。大型電子計算機の文字コードと本データの文字コードが異なる場合がある。以下に、JISで定められた以外の文字についてその字形と16進コードを示す。

文字	16進コード	文字	16進コード
①	8740	I	8754
②	8741	II	8755
③	8742	III	8756
④	8743	IV	8757
⑤	8744	V	8758
⑥	8745	VI	8759
⑦	8746	VII	875A
⑧	8747	VIII	875B
⑨	8748		

JISコードの記号である「◆」と「■」は、本データでは特殊な使い方をしている。

「◆」は、国立国語研究所の漢字コードで漢字テレタイプの盤面にはない文字を表すための記号である。それらの記号の後ろ2文字で、ある文字または記号を示す。たとえば、「約27◆○○C以上」の「◆○○」は温度を示す「°」である。

また、この調査では方程式等の式や数式は区別せず、1単位1語とした。「◆○式」と「◆○数」はそれを示す。

「■」は、漢字を表す。データ中の代表形と表記例の他の漢字でこれがどの漢字かわかる。ただし、次のレコードは■で示したが、片仮名のアイウを丸で囲んだ記号である。

試験管■■■	→	試験管①⑦
電圧E■■■, E■■■E■■■	→	電圧E⑦⑦⑦, E⑦①E①⑦
道■	→	道①
道■	→	道⑦

いずれも、正しくは報告書を参照されたい。

6.2 二つの調査の異同

「高校教科書の語彙調査」と「中学校教科書の語彙調査」の二つの調査は同じ調査方法で行ったが、別々に集計した別の調査である。したがって、代表形、判別情報の立て方に微妙な違いがあるかもしれない。注意ねがいたい。

6.3 訂正

『高校教科書の語彙調査』『五十音順M単位語彙表(付表)』で、数字の「1」が2レコードある。これは集計ミスなので、本データでは加算して1レコードとした。

また、本データファイルでは、報告書刊行後発見されたエラーを修正している。エラーの内容は、記号とされたり、間違って漢語とされていた英文字やギリシャ文字、元素記号他を外来語としたなど語種認定ミス他である。本データファイルの語彙量は、GOIRYO.DATファイルに示した。

7. データ使用上の注意

本データは、国立国語研究所が保有する言語研究用データを電子計算機が読める形で公開する刊行物である。これが言語研究の発展に寄与することを希望するとともに、この種の刊行を続けるために、以下の点についてご協力いただきたい。

- ①本データの利用に当たっては、研究所に著作権・版権があることをみとめ、その権利を侵害する行為は行わないこと。
- ②本データ作成のために点検を繰り返したが、まだエラーが残っているかもしれない。これらを完全になくすにはなお点検が必要だが、それでは公開が遅くなる。大量データの公開には常にこの種の問題が付きまとうが、利用者はこのことを承知の上でご利用願いたい。もしエラー等を発見された場合には、研究所に一報いただきたい。

- ③本データの利用による成果を公表する場合は、国立国語研究所の作成にかかるものであることを明示していただきたい。

8. 圧縮ファイルをもとに戻す方法

吉崎栄泰氏作成の高圧縮書庫管理プログラム LHA は、大きなファイルを圧縮して小さくしたり、これをもとに戻したりするプログラムである。その詳細は解説ファイル LHA.DOC を参照されたい。ここでは、2枚のフロッピーに入っている圧縮されたファイルをもとに戻す方法だけを説明する。

①1枚目のフロッピーをセットする。

②カレントドライブを①のフロッピーのドライブにする。

①のフロッピーがAドライブなら A: と入力し、リターンキーを押せば良い。画面の表示が、A:¥> となっていればカレントドライブがAであることを示す。

③次の通り入力し、リターンキーを押す。

LHA e もとに戻したいファイル もとに戻したファイルを入れるドライブ:
たとえば、KMG.LZH をもとに戻し、そのファイルをBドライブに入れる場合は次のとおり。

LHA e kmg b:

もし、ファイルを入れるドライブがフロッピーの場合、1メガバイトの空きエリアが必要なことがある。第3章のファイルの容量を参照されたい。

9. 謝 辞

高圧縮書庫管理プログラム LHA を使わせていただいた。作成者の吉崎栄泰氏に感謝の意を表する。

052759

1994年6月発行 [非売品]

Av
1
10020