

国立国語研究所学術情報リポジトリ

高校教科書用語調査の言語単位について

メタデータ	言語: Japanese 出版者: 公開日: 2017-06-13 キーワード (Ja): キーワード (En): 作成者: 麿岡, 昭夫, TSURUOKA, Akio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001302

高校教科書用語調査の言語単位について

鶴岡 昭夫

1. はじめに

現在国立国語研究所で行っている「高校教科書用語調査」は、昭和49年度に用いられた高校の理科および社会科の教科書、9教科（世界史・政治経済・人文地理B・日本史・倫理社会・生物I・化学I・地学I・物理I）について、最終修正が一応終わり、同語異語判別作業^(注1)にかかっている。この最終修正の終わった段階、すなわち前処理（プレディット）における問題がほぼ出つくし、その対策処理の終わった段階で、この用語調査で用いた言語単位（調査単位）の基本方針、単位切り作業手順、およびそれにもとづいて行われた単位切り作業で出現し、処理をした問題事項についてまとめて報告することにした。

なお、この調査研究にあたって、初期の段階で、担当の言語計量研究部の中に単位検討小委員会が置かれた。メンバーは石綿敏雄（当時第3研究室長、現在茨城大学教授）・野村雅昭（現在第2研究室長）・本稿筆者の3名であった。単位に関する基本方針および原則的項目についてはこの小委員会で試案を作成し、それを同研究部の全体会議にかけて討論し決定したものである^(注2)。また、実際の運用にあたって生じた問題については、単位切り作業を受け持った第1研究室（室長・土屋信一、室員・中野洋、同・本稿筆者）の中で合議の上決定した^(注3)。

したがって、本稿は、終始単位切り問題の検討に参画していたという関係上、筆者が言語計量研究部内の全体の検討結果をとりまとめて報告するものである、ということを付記しておく。

（注1）土屋信一「高校教科書の同語異語判別システム」（国立国語研究所報告61『電子計算による国語研究Ⅸ』1978年）

(注2) 鶴岡昭夫「高校教科書用語調査における言語単位」(国立国語研究所研究発表要旨集『用語用字調査と機械処理』1976年)

(注3) 細目については、作業者には作業規則書の追加版を渡し、また言語計量研究部内へは、部内報『季報』等により、逐次報告してきた。

2. 用語調査と高校教科書

2-1 これまでの用語調査

これまで国立国語研究所では、国語の基本語彙の設定、現代語の正書法の確定、標準語の確立をめざして、新聞・雑誌を中心とした用語調査を行ってきた。それらは次のようなものである(カッコ内は使用した調査単位^(注4))。

(1)朝日新聞^(注5)昭和24年6月分全数、述べ24万語 (β' 単位…ほぼ β 単位と一致する)

(2)婦人雑誌^(注6)主婦の友・婦人生活 昭和25年分の1/6、述べ20万語 (α 単位)

(3)総合雑誌^(注7)改造・中央公論など13誌 昭和27年7月～29年6月分の1/40 述べ23万語 (β 単位)

(4)現代雑誌90誌^(注8)世界・週刊朝日など90誌 昭和31年分の1/230、述べ53万語 (β 単位)

(5)現代新聞^(注9)朝日新聞・毎日新聞・読売新聞 昭和41年分の1/60、述べ300万語(短単位)・200万語(長単位)

これらの用語調査は、はじめ手作業で行われた。しかし、調査規模を大きくしてゆく必要性が認識されると同時に、50万語を越す調査は人手では大変な労力のいることと80万、100万語の調査は人手ではとうてい不可能であるということが、雑誌90誌の調査などからはっきりした。そこで、数を数えたり、一定の順序に並べたりする機械的な仕事を電子計算機にやらせることになって、現代新聞3紙の調査から HITAC 3010 型計算機を用いて行われるようになったわけである。

その後、高速漢字プリンタが実用化され、漢字仮名交り文の出力が容易になった。そこで、電子計算機、高速漢字プリンタを用いたシステムが考えられるようになった。また、新聞、雑誌のほかに、文学作品の用語調査を手がけてみ

ようという意見が出され、また、日本の近代文学の2大文豪といわれる夏目漱石と森鷗外を用語の研究を望む声が多かったことから、「漱石・鷗外を用語研究」という研究が行われた(注10) (この研究では、C単位、L単位、S単位の3種類の単位が用いられた)。

手作業の調査では2種類の単位を使えば、単位切り、カード化以後、すべて2度ずつ行い、全く別の調査を2回やるのと同じ労力が必要となる。しかし、電子計算機を使った調査では、それに比べれば少ない労力で複数の単位が使える。新聞3紙の調査では長単位・短単位の2種類、漱石・鷗外を用語研究では前述のように3種類の単位が用いられている。

(注4) 本項でいう過去の調査の単位は、簡単にまとめるとつぎの3つになる。

(1)文節となるもの……C単位

(2)文節を自立語と付属語に分けたもの…… α 単位・長単位・L単位

(3)最小単位(現代語で意味を担う最小の言語単位)の1回(または0回)

結合したもの…… β 単位・短単位・S単位

各グループにあるものは、たがいにほぼ一致するが、完全に一致しているわけではない。各調査の目的や担当者の言語観などによって少しずつ違いがある。細目については注5～注10の各文献を参照。

(注5) 国立国語研究所資料集2『語彙調査』1952年。調査単位については、P4～P9参照。

(注6) 国立国語研究所報告4『婦人雑誌の用語』1953年。 α 単位については、P19～P25参照。

(注7) 国立国語研究所報告12・13『総合雑誌の用語 前編・後編』1957年・1958年 最小単位(本書では「語源単位」とよぶ)、結合、 β 単位については、後編P10～P23参照。

(注8) 国立国語研究所報告21・22・25『現代雑誌九十種の用語用字(1)(2)(3)』1962年 1963年、1964年。最小単位、 β 単位については(1)のP6～P24参照。

(注9) 国立国語研究所報告37・38・42・48『現代新聞の語彙調査Ⅰ・Ⅱ・Ⅲ・Ⅳ』1970年・1971年・1972年・1973年。長単位、短単位についてはⅠのP13～P23参照。

(注10) 鶴岡昭夫「国語研究のための索引作成システム」(国立国語研究所報告59『電子計算機による国語研究Ⅶ』1976年)。C単位、L単位、S単位についてはP7～P9参照。

2-2 今回の調査の対象と目的

これまでの用語調査については前項に述べた。新聞・雑誌の調査はそれぞれ

数回ずつ行って一応の決着がついたものと考えた（これで終了して今後行わないという訳ではもちろんない）。文学作品の索引作成と用語研究も一応の成果を得た。次期用語調査ということで、書き言葉、話し言葉にわたってさまざまなものが候補にあがったが、高校の教科書（理科・社会・数学——現在、数学については中断）を取りあげることになった。研究所の外部にもそれを望む声があったようである。

高校教科書用語調査は、もちろんこれまでに行われた各種の書き言葉調査の延長として、それらの目ざしたところを補い、またそれらの調査結果と比較して検討することが第1の目的となっている。それを高校教科書で行うのは、国民が義務教育より高い段階での一般教養として、各分野の専門知識を身につける時に必要となる用語の範囲と使用語の実態を明らかにすることをもう一つの目的としたからである。さらに、新聞の用語調査、漱石・鷗外の用語研究という、電子計算機処理による用語研究を重ねてきた、その間のハードウェア、ソフトウェアの進歩を取り入れ、より効率のよいシステムを作り、精確な調査を行うことも目的の一つとした。

電子計算機を用いると各種の分析が可能になるという利点はあるが、しかしこのような大量調査にいろいろの目的を持たせることは危険なことである。だから当初の目的を上のようにしぼったのである。そして、最後にあげた目的に関連して、この調査によって作成され保存される言語データのファイル（漢字プリンタ出力による KWIC 索引および磁気テープファイルなど）は、他のいろいろな調査にたえる精度を達成するという方針にした。そのために人手を大量に投入して、修正を重ねること、同語異語の判別を行うことなどをシステムの中に組み込んだ。これらの作業は高速漢字プリンタを使用して作成した作業用の台帳（KWIC 索引表、原文イメージ帳、清書イメージ帳など）を用いることにより、能率よく、しかも精確に行うことに成功したといえる（なお、高校教科書用語調査のシステムについては本報告書の土屋・中野論文「高校教科書語彙調査システムとその設計思想」参照）。

2-3 今回調査の単位の決定まで

これまでの調査に用いられた単位については2-1に述べたようにいろいろな

ものがあつたが、今回の調査にあたっては、以下に述べるような検討を加えた結果、専門用語や、日本語の語構成などを調査するための長い単位（われわれは Word にちなんでW単位と呼ぶ）と、基本語彙および各専門分野の文章を構成する語彙の調査などのための短い単位（われわれは Morpheme に近いものとしてM単位と呼ぶ）との、2種類の単位を用いることになった(注11)。

まず検討の第1は、前項に述べた調査の目的がよりよく達成されることである。これは言うまでもないことで、調査の最優先事項である。つぎに、大量の調査であるからには、作業に負担がかかりすぎると調査自体の命とりになりかねない。作業の負担を軽くすること、これが検討の第2点であつた。具体的に言えば、単位切り作業は誰が行っても迷わず早く行えるようなものであることと、単位の種類を少なくすることである。とくに単位の種類については、2-1で述べたように電子計算機を用いると、単位の種類数が増えてもそれに比例して手間が増大することはないのだが、手間が省けるのは主として配列・集計等の機極的処理の部分（これとて作業量としては少なくないのだが）であつて、入力までの前処理および校正・修正といった中間処理など、主として人手にたよる処理では単位の種類にほぼ比例して手間がかかる（これはこれまでの新聞調査や漱石・鷗外の用語研究などから得た経験である）。したがって単位の種類はやたらに増やすわけにはいかない。

第3に検討されたのは、これまでの調査に用いられた単位（ α 単位系= α 単位・長単位・L単位、 β 単位系= β 単位・短単位・S単位）を——そのうちのどれかに決めて——用いるかどうかという点とである。たしかに調査単位が同じであつたほうが比較は簡単でしかも正確にできるし、また、新たに単位規則を作る必要もない、という利点がある。しかし、すでにこれまでも上にあげたように幾種類もの単位が用いられていて比較が簡単にできるというメリットはうすい。しかも、今回の調査は前に述べたように、基本語と専門用語という両方の目的を有しているものであつて、そのうえ電子計算機のハード、ソフト両面の進歩によって複数の単位の処理が可能になっているのであるから、それぞれの目的をはっきりさせて、新たに長短2種類の単位を考えなおしたほうがよいという結論に達した(注12)。

以上のようにして、本項冒頭に述べたように、W単位とM単位の2種類の単位を用いることになったのである。

(注11) 実際には助辞(26ページ8行目)に〈J〉という情報を付すことになったので、これを前のW単位語につなげた単位がいわゆる文節と等しくなり(2-1でいうC単位に相当)、3種類の単位が得られたことになる。

(注12) これまでの調査単位は、単一の単位のもとでの調査——基本語彙を中心としたもの——ということから、長い単位(α 単位系)は短かめにし、短い単位(β 単位系)は長めにして、両者が近づいているように感じられた。前回の現代新聞の語彙調査および漱石・鷗外の用語研究での長単位・L(Long)単位、短単位・S(Short)単位になって長い単位と短い単位という意識が生じたものと思われるが、これらも基本的には α 単位、 β 単位の方針を踏襲しているように考えられる。今回は、最初から2種の単位を目的別に設計しようとしたわけである。

3. W単位について

3-1 W単位の基本概念

前節で述べたように、W単位は専門用語や語構成などを調べるための単位である。その要求を満たすためには「第一次世界大戦」や「二酸化炭素」などのような長い調査単位にする必要がある。しかし、「第一次世界大戦」を1単位とするなら、「承久の変」も同じ程度の1単位として取りたいような気がするがどうしたものか、などという問題が出てくる。そこで次のような原則をW単位について定めた。

- (1) W単位は、修飾・並列・接続・中止・独立などの構文上の機能を持った最小の要素(一般に文節といわれる)の範囲をこえないものとする。体言の連体修飾は原則として認めない、すなわち名詞連続は切らないで全体を下の(2)での自立部分として扱う。また、いわゆる文節をこえないのであるから、上の「承久の変」は「承久の」と「変」とに分けて考える。「応仁の乱後」「高校を卒業後」なども、「応仁の」と「乱後」、「高校を」と「卒業後」に分けて考えるが、「十分の一税」のように「十分の」と「一税」とに分けられないものはこのまま例外的に残した。それと、「板の間」「茶の湯」など「～の～」の形で1体言に相当すると考えられるものは岩波国語辞典(第二版)の見出し語で名詞扱いにしているものを集めた一覧表をもとにして、その表にあれば全体を1体言として切らずに残した(注13)。

(2) 独立の意味を担い、形の上でも自立しているものを1W単位とする。非活用語の代名詞・名詞(いわゆる形容動詞語幹を含む)・副詞・連体詞・接続詞・感動詞などはそのまま1W単位となる(各品詞の認定は細則による)。活用語(動詞・形容詞およびそれらの終止連体形以外の活用形にいわゆる助動詞の付いたもの)は、末尾が終止連体形であるもの、命令形であるもの、および中止用法・修飾用法である連用形であるものが、1W単位となる。

(3) 上の(2)で述べた自立形態(W単位)の後に付いて、構文的なレベルで他と結びつけ、または述語の形を整える働きをするもの(実質概念を表さずに関係概念を表すもの)、これを我々は助辞と呼んでいるが、これは1要素(1助辞)ごとに1W単位とした(48ページ「助辞リスト」参照)。

(注13) 「～の～」の形をしたものは、岩波国語辞典に275語あった。そのうち、高校教科書に表れたのは、「天の川」「有りのまま」「絵の具(『水彩絵の具』の形で)」「かものはし」「茶の湯」「蜂の巣」「ひかげのかずら」「まのあたり」「身の回り」「もののあはれ」「山の手」「世の中」だけであった。なお、「～が～」の形の体言句はすべて1体言として扱った(例えば「おのがひとり子」「おのが民族」など)。

3-2 W単位の切り方

W単位は、実際の文章(今回の調査では高校教科書)に赤鉛筆で/を入れて切り分けるものである。その作業は、前項の原則および次のような細則によって行われた(細則に用いた例は本稿をなすに当たり、実際に出現したより適切なものに置きかえたものも一部ある)。

(1)記号は1単位とする。

例 /「/不戦条約/」/(/ケロッグブリアン条約/) /が/、/

〈例外1〉次の①②の記号は無視する。

①数字連続の中に現れる小数点・位取りカンマ。

例 /3.1514/ /2,600円/

②それがないときに全体がW単位(以下に規定する)となるようなものの中に現れる、次のような記号。

例 /アジア・アフリカグループ/ /P.T.A.会費/

／小・中・高教員／ ／ジョン＝F＝ケネディ／

〈例外2〉文中にあって、語と同じ働きをする記号（アルファベットの記号も含む）・記号連続およびそれらに語・数字の付いたものは全体を1W単位とする。

④／PTA／ ／ICBM／ ／O²／ ／H²O／ ／5 cm／
／WHO機構／ ／3000～4000℃／ ／（ ）／（～に入れる）

(2)助辞は1W単位とする（26ページの(3)、48ページの助辞リスト参照）

④／花／が／咲く／。／ ／イギリス／で／は／ ／争い／が／続いて／も／，／ ／学習／を／した／ ／外国／へ／行く／。／

《補》「～だ」「～です」の形をとるいわゆる断定助動詞・形容動詞語尾（その活用したもの、およびそれらに「う」「た」「て」「ば」などの付いたもの全体）は活用する助辞として切り離し（助辞リストF 51ページを参照）、全体で1W単位とする。

④／明らか／に／ ／おも／に／ ／容易／な／ ／静か／だ
ろう／ ／容易／だった／ ／私／だった／ ／私／です／

〈例外1〉「～の～」の形の体言相当句が、岩波国語辞典（第二版）に名詞扱いで立項されている場合は、この「の」を助辞とはしない（「の」の前も後も切らない。25ページ28行目参照）。

④／世の中／は／ ／身の上／を／

〈例外2〉「～が～」の形の体言相当句内にある「が」は助辞とはしない（「が」の前後を切らない）。

④／おのがひとり子／ ／おのが民族／ ／君が代／

〈例外3〉「この」「その」「あの」「どの」の「の」は助辞としない（切り離さないで全体を連体詞とする）。

④／この／時／に／ ／どの／よう／に／

《補》体言相当句の中部分に現れる「この」「その」などはW単位とせず
に、全体を体言扱いする。

④／電子そのもの／ ／大陸風そのまま／の／ ／春このかた／

〈例外4〉人名・地名で全体の内部に、助辞リストにある語を含むものは切

り離さずに、3・Fに従って処理する。

④／やまのうへの憶良／　／牧の原／　／高天が原／

〈例外5〉助辞リストにある語（例で下線を付けた語）と、それを除いた部分との結合が固定的であると認められる、つぎのようなものは切り離さずに全体を1単位とする。

④／あくまで／(cf.／あきる／まで／)　／あまりに／　／新たに／
／あるいは／　／いかが／　／いかに／　／いかなる／　／い
かん／　／いたずらに／(～人心をまどわし)　／いちがいに／
／一様に／　／一挙に／　／いっこうに／　／一手に／(～引き
うける)／一般に／(総じて、全般に、の意)／いやしくも／　／い
わんや／　／大いなる／　／大いに／　／おのずから／　／お
のずと／　／かならずしも／　／かりに／　／きつと／　／け
れども／　／現に／　／交互に／　／殊に／　／こまかに／
／さらに／(cf.／今さら／に／)　／しかも／　／しきりに／
／しだいに／　／実に／　／徐徐に／　／真に／　／少なくと
も／　／ずつと／　／すでに／　／絶対に／　／早急に／
／相互に／　／俗に／　／それとも／　／だが／(文頭で)
／互に／　／ただちに／　／ただに／　／単なる／　／単に
／ついに／　／では／(文頭で)　／でも／(文頭で)　／とかく／
／特に／　／ところが／(文頭で)　／ところで／(文頭で)　／とに
かく／　／ともに／　／とんでもない／(cf.／とんだ／)
／ならびに／　／なるべく／(cf.／なる／べく／して／なる／)
／にわかに／　／はるかに／　／ひいては／　／非常に／
／ひそかに／　／ひとえに／　／ふいに／　／まさに／　／ま
たは／　／間にあう／　／まもなく／　／みだりに／　／もつ
と／　／最も／　／やつと／　／よほど／　／わざと／
／割に／

(3)助辞を伴わない文節（修飾・並列・接続・中止・独立など、構文機能を有するもの）は一単位とする。問題となる点は下のA～Jに従って処理する。

それでもまだ問題がある（切るかどうかわからない）場合は、切らずに残しておく。（次に、修飾、並列、接続、中止、独立などの例を挙げる）

㊤／新しい／議会／ ／結局／わからなかった／ ／美しく／飾る／
／少ない／回数／ ／より高める／ ／二人／三人／と／集まる／
／よく／走り／よく／跳ぶ／選手／ ／美しく／青き／ドナウ／
／勝って／も／おごらず／ ／わかったら／答えよ／ ／しかし／わ
からなかった／ ／立ち止まって／考える／ ／立ち上がり／、／見
わたす／ ／汝／忘る／こと／なかれ／。／ ／毛沢東／万歳／

A. 体言的な形態をしたものに、形式的な用言（その転成名詞をふくむ）が直接続く場合は、間で切らない。

㊤／信用ある／ ／戦力ある／ ／心配なく／ ／むだなく／
／いかんなく／ ／ちがいない／ ／涙なし／に／ ／能率よく／
／表面近く／に／ ／出発する／ ／区別できる／

《補》 体言的な形態をしたものに、実質的な用言が直接続く場合は、体言的形態のあとを切る。

㊤／生涯／かけて／ ／板垣／死す／とも／自由／は／死せじ／

〈例外１〉とき・数を表す、体言的な形態をしたものが連用修飾語（主語も含む）となっている場合は、その体言的形態のあとで切る。

㊤（会合が）／今日／ある／ ／三回／あった／ ／二度／ない／

〈例外２〉体言的な形態をしたものが、連体修飾語によって修飾されている場合は、その体言的形態のあとで切る。

㊤／休む／暇／なく／ ／強大／な／戦力／ある／国／

B. 体言が連続していて、その一部分が連体修飾語をうけている場合は、その部分のあとを切る。

㊤／日本／の／首都／東京／ ／地区／の／大会／終了後／

《補》 ただし、次のような接尾的要素の付いた場合は切り離さない。

{ 中・上・内・外・間・あたり・後・前・的・視・ごと・以上・以内 }
{ ・以来・以下・同士・いかん }

例／その／円周上／　／同じ／時間あたり／　／同じ／範囲内／
／応仁／の／乱後／　／大化／の／改新以後／　／対岸／の／
火事視／　／戦う／国どうし／の／

C. 形容詞連用形（「～く」の形）と「なる」「する」「ない」との間は切り離す。

例／面白く／なる／　／高く／する／　／美しく／ない／

〈例外〉「なくなる」「なくする」は切り離さない。

例／事故／が／なくなる／　／大統領／が／亡くなる／　／犯罪／
を／無くして／　／亡くなった／人々／

D. 副詞に「する」が直接続く場合、間で切らない。「できる・なさる・いたす」など、「する」と近い構文的意味を持つものもこれに準ずる。

例／びっくりする／　／はっきりした／　／ゆっくりなさって／

〈例外1〉「する」および「できる・いたす・なさる」などが、「行う・やる」または「行える・やれる」に置き換えることのできる場合は、副詞との間を切る。

例／仕事／を／全く／しない／　／勉強／の／ほとんど／できない／
室／　／宿題／を／ゆっくり／する／

〈例外2〉「こう・そう・ああ・どう」に直接続いた「する」および「できる・いたす・なさる」等は切り離す。

例／そう／する／と／　／そう／して／　／そう／した／話／

E. 「こう・そう・ああ・どう」のあとに「いう」が直接続く場合、その間を切らない。

例／こういう／話／　／そういった／類／　／どういう／人／

〈例外〉「いう」が明らかに「話す・しゃべる」の意である場合は「いう」の前で切る。

例／大声／で／そう／いった／。／

F. 人名（姓および名。あだ名・しこ名・俗称・略称などもこれに準ずる）
・地名（行政区画名・地形名）・国名およびこれらを含む体言的連結は1
W単位とする。

④／本居宣長／　／徳川家康／　／後醍醐天皇／　／ルイ十四世／
 ／ブルボン王朝／　／吉田茂内閣／　／ワシントン大統領／
 ／海南島／　／江華島事件／　／鎌倉公方／　／バルカン半島／
 ／ドーヴァー海峡／　／ヨーロッパ大陸／　／富士山／　／横浜
 港／　／アメリカ／　／アメリカ合衆国／　／中国共産党／
 ／朝鮮民主主義人民共和国／　／ドイツ・イタリア連合軍／（→G）

〈例外1〉人名の前に来る肩書きは切り離す。二つ以上あるものはそれぞれに分ける。

④／陸軍大臣／東条英機／　／アメリカ大統領／ルーズベルト／
 ／指導者／レーニン／　／哲学者／プラトン／　／ドイツ人／シ
 ーボルト／　／前首相／佐藤栄作／　／故アメリカ大統領／ケネ
 ディ／（cf.／故ケネディ大統領／）　／征夷大將軍／権大納言／右
 近衛府大将／源頼朝／

〈例外2〉地名・国名などの連続したもので、次のように順に広い（また
 は狭い）ものになる場合は切り離す。

④／東京都／北区／西ヶ丘／　／アメリカ／アラバマ州／　／伊豆
 ／大島／　／東京／築地／の／

G. 並列は次のように切る。

④／走る／投げる／打つ／の／三拍子／　／A／B／および／C／
 ／日本／ドイツ／イタリア／対／連合国／　／首相／兼／文相／
 ／多芸／多能／な／人／　／素朴／自然／な／人がら／　／丸く／
 大きな／玉／

《補》並列ではない次のようなものは切らない。

④／酸素 O^2 ／を／　／港町横浜／（cf.／日本／の／主都／東京／）
 ／定数 n ／　／表A／　／表 $1-a$ ／

〈例外1〉一字漢語（漢字国名も含む）の並ぶ場合は切らない。

④／大小／の／　／真善美／を／　／東西南北／へ／　／日独伊
 ／の／三国／は／

〈例外2〉並列が、体言形態の中の一部で行われている場合は切らない。

⑧ / ドイツ・イタリア連合軍 / / アジア・アフリカグループ /

⑨ / 東京新大阪間 / / 酸素水素混合気体 /

H. つぎのような名詞が連続する場合、切り離してそれぞれを1単位とする

⑩ ①数についての単位が変わる場合。

⑪ ② / 3 m / 25 cm / / 1957年 / 1月 / 13日 / 午前8時 /

⑫ ③官公署・会社・学校、および部局などの名称

⑬ ④ / 大蔵省 / 造幣局 / / 国立国語研究所 / 言語計量研究部 / 第一研究室 /

I. 動植物名およびそれを含む体言句は1単位とする（27ページの(2)〈例外1〉のもの以外でも動植物名にはこれが適用される）。

⑭ ⑤ / ユキノシタ / / ワレモコウ / / ヒカゲノカズラ科 /

J. つぎの慣用句、およびそれを含む体言句は全体で1単位とする。

⑮ ⑥ / 人でなし / / ろくでなし / / 知らん顔 / / お気に入り /
 / 食わずぎらい / / 負けずぎらい / / よもすがら / / とにかく /
 / 間もなく / (cf. / 休む / 間 / も / なく / →29ページ23行参照)

以上がW単位の細則である。

4. M単位について

4-1 M単位の基本概念

M単位は、基本語彙、専門分野の文章を構成する語彙構造、語構成（W単位を構成するM単位という点から）などの調査を行うための短い単位である。短い単位ということでβ単位系のものをもとにして検討を加えた。その結果、短い方の単位は、和語・外来語については現代雑誌九十種の調査でいう「最小単位（現代語で意味を担う最小の言語単位）」を用い、漢語の場合だけは1回結合までを1単位（単独で用いられた1字漢語、および他のM単位に順次結合している1字漢語は1字で1単位）とするということを基本方針とした。

これは、 β 単位系のものでは和語の「母親」「行き過ぎる」や、外来語の「カラスライド」「ビヤホール」、混種語の「ビヤ樽」などの語（ β 単位語）から、それを構成する基本的な語と考えられる「母」「親」「行き（行く）」「過ぎる」「カラー」「スライド」「ビヤ」「ホール」「樽」などの語が得られないので、基本語の調査としては不十分ではないかと思われたこと、一方漢語については、漢字（音漢字）1 字で独立した漢語として用いられるよりも、2 字の結合で用いられることの方が圧倒的に多く、それだけ音漢字 1 字を語とするのには抵抗が感じられたことによる。

β 単位、短単位では、原則として最小単位の 1 回結合を 1 単位とし、接頭・接尾要素は 1 語 1 単位としていたが、M 単位では漢語だけを 1 回結合＝1 単位として、それ以外の和語・外来語および混種語中の和語・外来語部分では最小単位そのものを 1 単位としたところに違いがある。また、そういうことにした結果、M 単位では、 β 単位系で設けていた接辞的要素の規定が実質的に不要になり、作業への負担も軽減されるというメリットも生じた。

このように、M 単位がこれまでの β 単位系よりも短くなったのは「基本語」のとらえ方の変化ということも言えようが、やはり長短 2 種類の単位が使えることによって、短い方の単位で「母」と「親」、長い方の単位で「母親」（そのほか母親学級、母親方などの形も出て来る）という語形が得られるというシステムの変化——方法論の変化によるところが大きいのと考えられる。

4-2 M 単位の切り方

M 単位は、前節で述べた W 単位に分割するのが済んだ文章（赤線／の入ったもの）を用い、W 単位の中にある M 単位の切れ目に黒鉛筆で／を入れて切ったものである。本来、W 単位は文→W 単位、M 単位は文字→形態素→M 単位のような段階で処理が行われるものであるが、W 単位をこえる長さの M 単位はないものと考えて同一文章で二つの単位切りを行ったのである。この結果、文の切れ目の示されたもので M 単位を切り出す作業になり、単なる文字連続（文章）から M 単位を切り出すのよりも、作業者の負担は軽くなった。文字どうしの結合の組み合わせがはるかに少なくなったからである。

なお、この調査では、原文をできるだけ損なわないよう心掛けたが(註14)、ペ

ージの切れ目にあつて、両方にまたがるM単位は、後のページにある部分を前ページの末尾に書き移し、後のページの部分は抹消した。この調査のシステムがページ単位になっているため、M単位が2ページにわたっていると不便だからである（上のように書きかえても調査結果にはまず影響はない）。

M単位の作業は次の細則によって行われた（用例中、W単位の切れ目を／、M単位の切れ目を／で示す）。

（一）記号（W単位では無視したものも、ここでは含める）は1M単位とする。原子記号は1原子につき1M単位とする。

①／3／・／5／／ジョン／＝／F／＝／ケネディ／／（／）／内／の／
 〈例外〉数式・化学式・根号内の累乗根、および分母・分子ともに整数であるもの以外の分数は、全体を1単位とする（注15）。

②／ $a=3$ ／ ／ $a=\frac{x+n}{2}$ ／ ／ $2H^2+O^2 \rightarrow 2H O$ ／
 ／2／ \sqrt{x} ／ n ／個／の／ ／ $2\sqrt{x}n=ab$ ／ ／ $\frac{a}{b}$ ／ ／ $\frac{b}{a}$ ／
 x ／の／ ／三角形の面積＝底辺×高さ× $\frac{1}{2}$ ／

《補》整数を分母、分子に持った分数が、文中で語と同じ働きをのっている場合は「分子／分母」（／はW単位の切れ目ではなく、1/2、3/4などの斜線の意味）として表す。

③／国土／の／1／／／2／は／

（二）記号以外の語は、次の(a)～(d)の最小単位をもとに(1)～(3)の規則によりM単位に分割する。数（アラビア数字、漢数字、ローマ数字とその連続）を含むものはさらに（三）によって処理する。

〔最小単位の例〕

(a)固有名詞の最小単位の例

1) 人名

姓…藤原 源 山田 呉 孫 ケネディ

名…定家 義経 一郎 三桂 文 ジョン

通称・雅号・芸名・しこ名・あだ名など……徳球 エノケン 古今亭

十返舎 一九 竹本 義太夫 双葉山 石部 金吉

2) 地名

国名・行政区画名・地域名・地形名などでは、「国・王国・共和国・合衆国・連邦・都・道・府・県・市・町・村・街・通り・山・川・島・海・湾・海峡・岬」など、類概念を除いた部分を最小1単位とする。

④日本(国) ヨルダン(王国) ソヴィエト(連邦) 東京(都)
富士(山) テネシー(川・州) ミッドウェー(島)
ハワイ(諸島・州) 太平(洋) ヨーロッパ(州) 唐

(b)和語の最小単位

④手[て, た] 雨[あめ, あま, さめ] 物(もの) こと
美しい[美しく, 美しかろ, 美しかっ……] 行く[行か・行き……]
きつと 最も (出発)する (行け)ば (読も)う
(見)た (行っ)て (書か)ない (知ら)なかつ・た
(行っ)たろ・う (面白)がる (面白)そう お(父)さん
(美し)さ (重)み (寒)げ (丸)める・(丸)まる
(注・「める」と「まる」でセットになるもの) (ふた) つ(ひと)り
(とお)か (茶)の(間)

(c) 外来語(和製外来語も含める)の最小単位

1) 一般に原語での1最小単位

④ビーカー ガラス インフレーション アルミニウム カメラ
ミシン マス コミュニケーション メス シリンダー ガス
バーナー フェノール フタレイン メー デー クロル カルキ
バクテリア クロロ フィル キサント プロテイン

2) 日本語では分解不可能な形に熟合したもの

④クーデター ヘクター(ただし ha は h と a とに分ける)

ネクタイ ハンカチーフ ハンカチ

3) 漢字をあてたもの

④倶楽部 背広 麦酒[ビール] 煙草[タバコ] 燐寸[マッチ]
型録[カタログ]

4) ローマ字・カナで略したもの

(d)漢語の最小単位・(和製の音漢字も含む)の例。漢語の認定は新潮国語辞典(改定版)による。

例一 二 三 愛 働 義 信 用 的 風

[M単位の例]……次の(1)(2)(3)による。

(1)漢語

漢語要素(上の(d))だけからなるW単位内では、最小単位の1回結合したものを1M単位とする。W単位の中で、このM単位の前または後に順次に結合した最小単位・最小単位1回結合は、それぞれ1M単位とする。結合とは、最小単位どうしの意味的なつながりで、その順番により1回結合、2回結合……となっていくものである。

(結合例)

1 回結合例



/人間/

/愛情/

/上下/

2 回結合例



/公務員/

/外交官/

/化合物/



/大漁場/

/貴金属/

/濃硫酸/



/国会開設/

/天気予報/

/硫酸亜鉛/

3 回結合例 (一部)



/植民地化/

/先進国間/

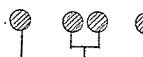
/試験管内/



/大企業間/

/短周期性/

/強塩基性/



/新歴史劇/

/強塩基性/

/凝固点降下/



/不況下倒産/

/化学的性質/

/凝固点降下/

以上は結合の一例であるが、他の場合も同様にして考える。

例 /超高速/ /現副首相/ /洪積世人類/

/悪人正機説/ /次亜塩素酸/ /安全保障条約/

/国際連盟脱退/ /個々人/ /公民権法改正/

／共有／結合／結晶／ ／塩基／性／酸化／物／ ／労働／者／
 代表／会議／ ／国会／開設／期成／同盟／ ／急進／的／民族／
 主義／者／

〈例外1〉 3つ以上の漢語要素の並列からなるW単位では、それぞれの要素を1M単位とする。

④／真／善／美／の／ ／衣／食／住／を／ ／都／道／府／県／

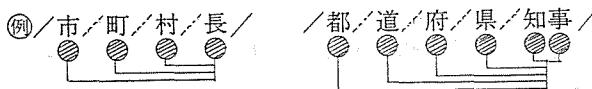
《補》2つの漢語要素の並列は1回結合として1M単位にする。

④／町村／の／ ／大小／の／

〈例外2〉 2つ以上の漢語要素が対等の資格（並列）で、他の要素（またはその結合）と結合している場合、次のように処理する。



〈例外3〉 3つ以上の漢語要素が並列で他と結合している場合、それぞれを1M単位とする。



〈例外4〉 3字以上の漢語でも、現代語として、切り離せないと考えられる場合は切らずに全体を1単位とする。

④／阿弥陀／ ／過不足／ ／加地子／ ／勘解由／ ／義太夫／
 ／下剋上／ ／三味線／ ／浄瑠璃／ ／二十日／
 ／雾囲気／ ／曼荼羅／ ／未亡人／ ／官公労／ ／不可思議／
 ／不思議／ ／有耶無耶／

〈例外5〉 以下のように、切り方に問題のある（／○○／○／か／○／○／○／か決定しがたい場合）は切らない。

⑨／一部分／ ／大部分／ ／海水浴／ ／加速度／ ／過半数／
 数／ ／下半部／ ／下流域／ ／原住民／ ／原動力／
 ／皇太子／ ／極限界／ ／後半期／ ／地表面／ ／水産業／
 ／生化学／ ／小農民／ ／諸国民／ ／電磁気／
 ／農産物／ ／不可欠／ ／副産物／ ／暴風雨／

(2) 漢語を含む混種語

漢語要素(d)と、(a)・(b)・(c)の少なくとも1つとの合成からなる混種語は、(1)と同様に、1回結合を基本としたものに分割する。1要素(1最小単位)が単独で切り出されたものはそのまま1M単位となるが、1回結合には次の三とおりが考えられる。すなわち、イ) 漢語要素だけの1回結合、ロ) 漢語要素と他の要素 ((a), (b)または(c)) との1回結合、ハ) 漢語要素以外の要素 ((a), (b)または(c)) の1回結合である。この1回結合は、それぞれ次のように処理する。

イ) 漢語要素の1回結合が残った場合、その1回結合を1M単位とする。

⑩／アレルギー／鼻炎／ ／草／野球／

ロ) 漢語要素と他の要素 ((a), (b)または(c)) との1回結合は切り離してそれぞれを1M単位とする。

⑪／アレルギー／性／ ／イオン／化／傾向／ ／生田／流／

／スペイン／風／ ／フランス／人／ ／ソ／連／ ／セ／氏／
 ／訪／ソ／ ／塩／ビ／ ／実／に／ ／暗／に／

《補1》 次のような重箱よみ、湯桶よみのものに注意。

⑫／絵／心／ ／本／箱／ ／額／縁／ ／場／面／ ／場／
 所／ ／手／本／

《補2》 漢語の固有名詞どうしの結合は(3)の3 (42ページ) 参照。

《補3》 助辞リストにある語と漢語要素との結合については(3)の1の《補5》 (40ページ) 参照。

〈例外1〉 1字漢語と1字漢字の国名・地域名(略語も含む)との結合は切り離さない。

⑨ / 米 国 / / 漢 字 / / 楚 歌 / / 訪 米 / / 訪 欧 /

〈例外2〉 姓などを臨時に音読したものは、固有名詞要素とはしない。

⑨ / 平 家 / / 源 氏 / / 藤 家 / / 菅 家 /

(cf. / 宗 / 氏 / / 李 / 氏 / / 孫 / 家 / ……実際の姓である)

〈例外3〉 1字漢字の中国人名に「子」の結合したものは切らずに全体を1M単位とする。

⑨ / 孔 子 / / 孟 子 / / 莊 子 / (cf. / 淮 南 / 子 /)

〈例外4〉 1字漢語に活用語尾やサ変動詞の付いたものは切らないで、全体を1M単位とする。

⑨ / 力 む / / 愛 する / / 信 じる / / 信 ずる /

(cf. / 恋 / する / / 値 / する / ……和語にスルが付いたもの)

ハ) 1回結合をしている部分が漢語要素を含まない場合はそれぞれを切り離してM単位(2つ)とする。

⑨ / 立 / 会 / 人 / / ピンク / 色 / / フランス / ドイツ / 国境 /

(3)漢語を含まない語(W単位), すなわち和語, 外来語, 固有名詞, およびそれらの混種語は, 1最小単位を1M単位とする。

1. 和語

⑨ / た か ら / も の / / ま / ご ころ / / わ れ / わ れ / / 手 / 続
き / / 支 / 払 い / / 試 / 合 / / 青 / 空 / / あ り / の /
ま ま / / 面 白 / が る / (cf. / 広 が る / ……拡大スルの場合)
/ 真 っ / 黒 / / 子 供 っ / ば い / / 人 / で / な し / / 相 / つ

い／で／ /目／深／〔まぶか〕 /赤／み／ /一／人／

《補1》 最小単位と考えられるものを含んでいても、それを切り離すと後に最小単位と思えないものがある場合は、切り離さずに全体を1M単位とする。

⑨ /けだもの／ (cf. /け／もの／) /まなこ／ /まぶた／
/まなじり／ /青い／ /くだもの／ /たなごころ／
/あるいは／ /めめしい／ /長々しい／ (cf. /長／長／と／)
/きっかけ／ /しからば／ /とかく／ /とにかく／
/とにもかくにも／ /ともかく／

《補2》 次のような、動詞を起源とする副詞・接続詞は、全体で1M単位とする。

⑩ /はたして／ /けっして／ /きわめて／ /すべて／
/はじめて／ (cf. /始め／て／) /例えば／ (cf. /例えれ／ば／)
/あくまで／ (cf. /あきる／まで／) /つとめて／ /あらためて／
/概して／ /かえって／ /ついで／ /かねて／ (～計画中の) /たえず／

《補3》 形容詞語幹に付く五段活用語尾「む」「ぶ」とその活用形およびその連用形転成名詞の末尾の「み」「び」は切り離さない。

⑪ /苦しむ／ /楽しむ／ /慈しみ／ /悲しび／

《補4》 和語の畳語はそれぞれ1M単位とする。

⑫ /あれ／これ／ /だれ／それ／ /人／々／ /それ／ぞれ／
/我／々／ /いや／いや／ /恐る／恐る／ /泣き／泣き／
/一／人／一／人／

《補5》 助辞リストにある語を構成要素にもつ次の語 (28ページ参照) は、助辞要素を切り離した部分が他にも用いられる場合は切り離す。そうではないもの (語幹に自立性がみられず、その形式での用法が固定しているもの) は、切り離さずに全体を1M単位とする。

(切り離すもの)

⑬ /あまり／に／ /新た／に／ /いか／に／ /いか／な

る／　／いたずら／に／　／いちがい／に／　／一様／に／
 ／一挙／に／　／いっこう／に／　／一手／に／（～引きうけ
 る）／一般／に／（全体・総体ニ、の意）　／いは／ん／や／
 ／いわば／　／かならず／しも／　／極度／に／（cf.／極度／に
 ／まで／達し／た／）／現／に／　／こまか／に／　／さら／
 に／　／しきり／に／　／しだい／に／　／じつ／に／
 ／徐々／に／　／真／に／　／少なく／とも／　／絶対／
 に／　／早急／に／　／相互／に／　／俗／に／　／だ／
 が／（文頭）　／互い／に／　／ただ／に／　／単／なる／
 ／単／に／　／つい／に／　／で／は／（文頭）　／で
 ／も／（文頭）／特／に／　／ところ／が／（文頭）　／とこ
 ろ／で／（文頭）　／とも／に／　／ならび／に／　／なる／
 べく／（cf.／なる／べく／し　て／なる／）　／にわか／に／
 ／はるか／に／　／非常／に／　／ひそか／に／　／ひとえ
 ／に／　／ふい／に／　／まさ／に／　／また／は／
 ／間／に／あう／　　／ま／も／なく／　／みだり／に／
 ／よ／ほど／　／わざ／と／　／割り／に／

（切り離さないものい）

㊦／いかが／　／いかん／　／いやくも／　／大いなる／
 ／大いに／　／おのずから／　／おのずと／　／きっと／
 ／けれども／　／ことに／（殊）　／しかも／　／ずっと／
 ／すでに／　／それとも／　／ただちに／　／とんだ／
 ／とんでもない／　／ひいては／　／もっと／　／最も／
 ／もっとも／（尤）　／やっと／

《補6》 現代語で一語と考えられるつぎのものは切らない。

㊦／あらゆる／　／いわゆる／　／みずうみ／　／報せる／
 ／知らず／（五段の使役動詞）　／つまずく／　／うなずく／
 ／生まれる／（ただし／子供／を／生ま／れる／……尊敬）／あたか

も／　／あらかじめ／　／いまだ／　／いまだに／　／いわ
く／　／大きな／　／同じく／　／仮名／　／くちばし／
／こちら／　／どちら／　／すなお／　／しまう／(読んで～)
／とらえる／　／とられる／　／なべ／　／とりわけ／
／はなはだ／　／はるばる／　／前もって／　／さまたげる／
／さまよう／　／ついたち／　／ついばむ／　／とおざかる／
／なかば／　／おける／(国内に～)　／ひきいる／　／望
ましい／　／やくざ／　／ふるまい／　／屋根／　／めざま
しい／(cf./目/覚める/)　／もてはやす／　／もはや／
／こんにちは／　／さようなら／　／こんばんは／　／ひとり／
(～日本のみならず)

2. 外来語

④／カラー／スライド／　／アルミニウム／イオン／　／オイル
／ダラー／　／フェノール／フタレイン／　／マス／プロ／　／キ
サント／フィル／　／キサント／プロティン／　／メタ／クロロ／
ジ／アミド／　／バイオロ／クロロ／フィル／

《補》 外来語，またはその略語に活用語尾の付いたものは全体で1M単位とする。

④／デモる／　／サボッ／た／

3. 固有名詞

①人名…姓と名はそれぞれ1M単位とする。

④／藤原／定家／　／徳川／家康／　／孫／文／　／李／鴻章／
／セオドア／ルーズベルト／　／ジョン／＝／F／＝／ケネディ／
／ナポレオン／三世／(三で／三／世／)　／十返舎／一九／
／豊竹／山城少掾／　／大鵬／幸喜／　／源／九郎／義経／(三
で／九／郎／)

《補》 ルビにのみあらわれる，姓・名間の「の」は無視する。

④／源^{みなもと}義経^{よしつね}／→／源／義経／(／みなもと／よしつね／)

②地名（類概念は外す → 35ページ参照）

㊦／日本／海溝／ ／東京／湾／

《補》 漢字一字の国名（略語も含む）の並列，結合は切り離す。

㊦／日／中／ ／米／英／ ／薩／長／

〈例外〉類概念を外したものが，地名，地形名として独立しないものは切らない。

㊦／荒川／ ／大島／ ／黄河／（cf.／揚子／江／） ／漢江／

4. 和語・外来語・固有名詞の混種語

㊦／赤／ランプ／ ／巻き／スカート／ ／セメント／がわら／
 ／山田／さん／ ／東京／寄り／ ／パリッ／子／ ／ケネディ
 ／ラウンド／ ／ヨーロッパ／スタイル／ ／メコン／・／デルタ／

㊦ 以上の規定で分けられたものの中に，数を表す要素が含まれていたら，その部分をさらに分けて，各数をそれぞれ1 M単位とする。数は，アラビア数字，ローマ数字・漢数字（これまで，漢語要素として扱ってきた）のほか，「何十」「数百」「幾回」「百余」などの「何」「数」「幾」「余」等も数として扱う。

㊦／第／三／回／ ／四／分／の／一／ ／四／五／回／ ／五／
 時／七／分／ ／3／. ／1／4／1／6／ ／1／9／0／6／
 年／ ／1／，／0／0／0／円／ ／三／カ条／（cf.／箇条／書
 き／） ／三／カ／国／ ／何／回／ ／一／^{いち}日^{にち}（cf.／^{ついたち}一日／）
 ／五／百／十／五／人／ ／数／十／万／人／

〈例外〉その数だけで特別の意味であって，数量を表すことに主眼のない次のようなものは，数の部分を切らない。

㊦／一員／ ／一因／ ／一円／（／一円／的／） ／一応／
 ／一概／（／一概／に／） ／一群／（cf.／第／1／群／） ／一時／
 （／一時／金／，／一時／的／，cf.／一／時／十／分／） ／一族／
 （藤原～） ／一大／（～文学集） ／一团／ ／一段／（／一
 段／と／） ／一堂／ ／一人／（／一人／前／の／大人／，cf.
 ／一／人／前／の／寿司／） ／一番／（～よい） ／一部／（全

部ニ対して) /一部分/ /一瞥/ /一面/ (～において,
 cf. /一/面/鏡/) /一門/ /一躍/ /一樣/ (/一樣
 /に/) /一率/ /一連/ (～の事件)
 /一路/ /一貫/ (/一慣/性/) /一挙/ (/一挙/
 に/) /一見/ (/一見/し/て/) /一国/ (/一国/内
 /) /一種/ (～の) /一生/ /一心/ /一身/
 /一進/一退/ /一世/ (～を風靡し) /一体/ (/一体/
 化/, cf. /一/体/の/仏像/) /一旦/ /一端/ /一手
 / (/一手/に/ひき/うける/) /一定/ /一転/ (/一
 転/し/て/) /一途/ /一派/ (/文学/界/一派/)
 /一变/ (/一变/する/) /一辺倒/
 /一方/ (増加する～) /五輪/ (オリンピックの意) /
 再三/ /十字/ (/十字/架/, /十字/路/, cf. /十/字/分
 /) /空ける/ /十分/ (～に調べる) /十両/ (スモウで)
 /千軍/万馬/ /千載/ (/千載/集/, cf. /千/載/一/遇/)
 /千差/万別/ /百姓/ /万感/ /万物/ /万一/
 /万葉/ (/万葉/集/) /唯一/

M単位の細則は以上のとおりである。

(注14) ここで述べる、2ページにわたるM単位のほか、明らかな誤植は訂正した。そのほか、W単位語(前節)の中間部分に割り込んだ注をW単位語の後へ移すようにしたが、これについては次節(45ページ)参照。

(注15) 高校教科書には数式が多く、すべてを用語調査の対象にすると、調査結果にかなりの影響が出てくる。そこで、数式は◇○数、◇○式などの記号にして扱うことにした。ただ、整数を分母・分子とする分数は、本文中によく現れ、また、ことばとしての機能をもつことが多いので、これは残すことにした。

5. 実際の単位切り作業

5-1 単位切り作業の実例

第4節、前節でのべた原則、および細則にしたがって実際の文章をW単位とM単位に切ると下のようになる。なお、前述のように、まず赤スラッシュ(本

稿では／)でW単位に分け、その中でM単位の切れ目に黒スラッシュ(本稿では／)を入れる作業であった。W単位は／から／までであるが、M単位の場合は、W単位内に／の区切りがなければW単位=M単位、／の区切りがあればその斜線によって分けられたそれぞれがM単位となる。

／ピルピン／酸／は／脱／炭酸／さ／れ／て／、／活性／化／さ／れ／た
／酢酸／に／なっ／た／のち／、／まず／オキサロ／酢酸／と／反応／し／
て／クエン／酸／と／なり／、／脱／水酸／酵素／の／はたらき／で／水素
／を／失い／、／脱／炭酸／酵素／の／はたらき／で／二／酸化／炭素／を
／失う／反応／を／くり／かえす／うち／、図／4／2／の／よう／に／オ
キサロ／酢酸／に／もどる／。／

／貨幣／は／もと／もと／金／など／の／貴／金属／で／作ら／れ／て／
い／た／。／そして／、／たとえば／1／0／円／の／金貨／は／、／地／
金／に／し／て／も／1／0／円／の／値／うち／を／持っ／て／い／た／
。／

5-2 出て来た問題

単位切り作業、および中間の KWIC 索引による校正作業といった、主として人手にたよる作業は、言語計量研究部第1研究室を中心に行われた。その作業でさまざまな問題点が発生した。その多くは、当初の予測を越えて、規則書に規定されていないケースであり、調査ではその都度補正された。本稿では、それらの点は一応整備された姿になっている。その中で、単位の本質とかかわる問題についてはここでふれる(それぞれの規則を参照するように示す)。また、上にあげた諸規則で処理をしてもなお問題になった点についてもここでふれる。

(1) W単位で切るかどうか

a) 語中の注釈

前節の注14(44ページ)に述べたように今回の調査では本文をできるだけ損なわないように心がけた。その方針で「1958(昭和33)年」という個所を、これまでのW単位規則で処理すると、／1958／(／昭和38／)／年／のようになってしまって、「1958」と「年」とがはなれてしまう。W単位語としては

「1958年」であるべき間にカッコの補注が割り込んだものと考えられる。そこで、そのような割り込んだ注は軽視して、W単位の切れ目までさがらせることにした（実際には「年」をカッコの前まで書き移すのだが）。すなわち／1958年／（／昭和33／）のようにしたのである。そのようにしたのは、ほかに「揚子（ヤンツー）江」→「揚子江（ヤンツー）」などがあった。

b) 「でも」か「で／も」か

「めしでも食おう」の「でも」は「で」と「も」に切れない。一方「学校でも教えている」の「でも」は「で」と「も」に切れる。このようにはっきりしていれば問題にならないのだが、実際の調査では作業者の迷うような場合が少なくなかった。そこで、「で」と「も」に分けるのは①「も」をとっても構文的に通用する、②対応する限定形「では」の形が構文的に通用する、③「であっても」の意味を持ち、これにおきかえられるもの、という規準を設け、①②を満たす、①③をみたす、②③をみたす、のいずれかの場合として、一応の統一をとるようにした。しかし、上の「通用する」というのは主観的な判断で、作業者によって、また同一作業者でも時間の経過とともに変わることがしばしばある、という問題が残った。この調査での「でも」「で」「も」については、一応の決着の付いたもの、とだけしかいえない。今後の課題とした。

c) 「～に」の形の連用修飾語

「静かに」の「に」はいわゆる形容動詞の語尾で、1W単位である。しかし、「非常に」の「に」には、形容動詞語尾とするだけ活用語尾がそろっていないものと考えられる。すなわち、この「に」は助辞とするよりも前の部分に固定されているということになる。そこで、W単位規則(2)の〈例外5〉28ページ参照)の規定の中に、「～に」の形で連用修飾語（副詞）として固定的に用いられるもののリストを作って入れた。ここにあげたのはこの調査で現れた範囲内であるから、他の調査にこの規則を用いる場合は適宜増補する必要がある。なお、「～に」の形をしていて「～」にの部分に他に用いることのないもの、例えば「すでに」「殊に」などについては、M単位の問題である（M単位規則(3)1の《補5》40ページ参照）。

〔2〕 M単位で切るかどうか

a) 2とおりに読めるもの

〔書き言葉の調査で問題となるものの1つに、同形異音同義（または類義）の語の処理である。例えば、「数」をカズと読むかスウと読むかという類である。読み方が2とおりあって1つに決定しがたい場合、今回の調査はその場の判断にまかせて（要するにどちらにしてもよいこと）、むりに統一をとらず、それらの語の調査は別に行うことにした。このような処理で、漢字1字の自立語の場合はM単位にはひびかないが、2字の漢字からなる語の場合は、それが漢語であるか和語であるかによって単位切りにまでひびいてくる。例えば、「牧場」「工場」をボクジョウ、コウジョウと読めば漢語で、1回結合であるからそれぞれ1M単位となるが、これをマキバ、コウバと読めば「まき」と「ば」とが和語だから、それぞれ／牧／場／、／工／場／のように切れる。これらもその都度の判断にまかせて、むりに統一はとらなかった。結果としてその多くは漢語となったようであるが、これらについての調査も、別に行われる予定である。

b) 漢字1字の訓が2最小単位以上からなるもの

漢字の訓の中には、「陥る」（落ち＋入る）や「卵」（玉＋子）など、2最小単位から成っていると考えられるものがある。M単位では、当用漢字音訓表にあるものは分割しないことにした。「卵」「玉子」「たまご」「陥る」「落ち入る」「おちいる」はそれぞれ1M単位である。また、表外字や表外音訓の場合は、漢字1字で書かれているもの、例えば「^{しろがね}銀」は1M単位、かなのもの、例えば「しろがね」は「しろ」「がね」の2M単位とすることにした。ただし、「一人」「二たび」に関しては、「独り」「再び」に「ひとり」「ふたび」の訓があるがそう書かれる場合と意味に違いがあるので、／一／人／、および／二／たび／と切ることにした。

c) 「町村長」「原水爆」などの漢語

これまでのβ単位では、2個の漢語要素が、並列で他の要素と結びついた「町村長」のような語は、「町長・村長」として2枚のカードにしていた。しかし、この調査では、原文を損わないようにすること、漢字の調査で、実際に

は1回の漢字が2例と計算されてしまうこと、W単位語として「町村長」という語がすでに認定されていることなどの理由で、三字漢語として残すことにした(M単位規則(二)(1)の〈例外2〉37ページ参照)。しかし、三字以上の並列の場合は、出現する形が限られているうえ、それぞれの漢字の独立性の強いものが多いので(注16)、それぞればらばらにすることにした(同じく〈例外3〉37ページ参照)。

(注16) 2字の並列の場合は「原水爆」のように、「原水の爆」などと切り離して言えないものが少なくないが、3字以上の並列の場合は、「市町村の長」「都道府県の知事」のような言い方ができることが多い。

6. おわりに

W単位とM単位は、まだ高校教科書調査を経ただけで、まだ十分に完成されたものとは言えない。しかし、このように大量の言語データを、構文論的な語と形態論的な語に分離して、しかもほぼ同時に処理してゆくということについてはこの調査が最初の試みであるといえよう。これは、自慢めいてしまうが構成メンバーである各研究員、研究補助員およびこれを助けた多くのアルバイト達の協力と、ソフトウェア、ハードウェア両面にわたる電子計算機の進歩とのたまものである。今後、このW単位、M単位がいろいろな調査を経て発展してゆくことができればさいわいである。

付 助辞のリスト

A いわゆる格助詞

1. が……………花が咲いた。 私が読んだ本。
 ～を見るがごとく。 ～がゆえに
2. から……………今月から始める。 学校から帰る。
3. って……………遅いって言われる。
4. で……………学校で習う。 風邪で休む。 新聞で知る。
5. と……………友達と行く。 だめだと思う。
 ありとあらゆる。 生きとし生けるもの。
6. に……………学校に行く。 風邪にかかる。
7. の……………黒のスーツ。 私の読んだ本。
 これは私のだ。

生まれながらの～。 しかしながら。

18. など……………飯など食べる。 アメリカなど六カ国訪問。
19. なり……………何なりと尝试てみる。 せめて便りなりよこせ。
20. のみ……………あとは祈るのみだ。
21. は……………私は少年です。 これは本だ。 行くには行く。
22. ばかり……………泣くばかりでない。 少しかばかりの金。
23. ほど……………死ぬほどの思い。 あせればあせるほどうまいかない。
24. まで……………君までそんな事をいう。
25. も……………今日も雨だ。 学校にも行かない。 見てもいい。
26. や……………それを見るや立ち上る。 するやいなや。
27. やら……………どうやらわかった。

D いわゆる接続助詞

1. が……………あるにはあるが小さい。 雨が降ろうが構わない。
2. から……………若いから元気だ。 出かけるから仕事する。
3. けれども……………雨が降るけれども行く。 知らないけど。
4. し……………子供じゃあるまいし。
5. って……………行ったって仕方がない。
6. で……………～を固執しないで。
7. と……………雨がやむと出掛ける。 何があろうとかまわない。
8. とて……………何が起ったとてかまわない。
9. とも……………板垣死すとも自由は死せじ。 雨が降ろうとも。
10. なり……………帰るなり大声でとなりつける。
11. に……………思うにこれは人災だ。 要するに頭がいいんだ。
12. ので……………雨が降るので外出しない。
13. のに……………知っているのに返事もしない。
14. も……………遅くも来月には返してほしい。
15. ものの……………そうは言ったものの。 ～とは思うものの。

E. いわゆる終助詞・間投助詞

1. い……………本当かい。 だめだい。
2. か……………本当か？ だめか。
3. かな……………ああ。短いかな人の生命よ。
4. けれども……………勿論それはそうだけれども。
5. さ……………わかっているさ。
6. ぜ……………だめだぜ。 たのんだぜ。

7. ぞ…………だめだぞ. いいぞいいぞ. なんの謂ぞやと.
8. って…………今日帰って来たんだって.
9. とも…………そうだとも.
10. な…………解ったな. ここへ近寄るな.
11. に…………さぞおもしろかったらうに.
12. ね…………面白いね.
13. の…………とっても面白いの. 来るの?
14. のに…………解っていれば注意したのに.
15. や…………面白いや.
16. やら…………今日は何があるやら.
17. よ…………そうよ. だめだよ.
18. わ…………変だわ. それには参ったわ.

F いわゆる助動詞（その活用形に、接辞『て、ば、う、た』の付いたものは全体で W単位とし、M単位で細分する。……例：／いい／だろ／う／）

1. ごとし…～のごとく. ～を見るがごとき.
 2. だ…………私だ. 立派な人. いいだろ(う). 静かである.
立派だっ(た). 主なる神. ～したなら(ば).
 3. たり…………治者たるにふさわしい. 堂々たる姿.
 4. です…………私です. 立派です. いいでしょ(う).
静かでし(た).
 5. べき…………行くべきだ. それをを考えておくべきだ.
考えるべから(ず).
 6. らしい…………男らしい. 雨が降るらしい. 男らしくっ(た).
- (注) 「だ」には、いわゆる断定の助動詞と、いわゆる形容動詞語尾（ようだ、そうだ、ふうだの活用部分を含む）を含む。