The method to discribe word in language data processing

言語情報処理における語の把握

土屋信一

1. はじめに

この稿は，語彙調査，とくに電子計算機を使った語彙調査において，「語」をいかに把握すべきであるかについて考えたものである。語彙調査において，何を「語」としてとらえ，分析するのかということは，自明のことと考えられてか，あまり論じられた形跡がない。国立国語研究所が創設以来一貫して実施してきた数回の語彙調査も，一二を除き，同様である。しかし，電子計算機を使った語彙調査では，言語を機械の中に送りこみ，何らかの処理を施して後，取り出すというプロセスを経るため，何をもって「言語」とするか，そしてそれをどう記述し，記号化するかということが，非常に重要な問題となってくる。電子計算機は「クズを入れればクズが出る」機械だと言われるが，これはデータの質の良否だけでなく，データ（この場合，言語）の記述の形式が十分に考慮されていなければ，良い分析結果が得られないことも意味していると考える。この稿では，言語の記述形式として，従来採られている，表記に使われる文字列をそのまま用いる方式の問題点を指摘し，代りに，「語列」と呼ぶべき形式を設定しようとするものである。この考え方は，今回の高校教科書の用語用字調査の同語異語判別処理で活かされている。と言うよりも，同語異語判別処理において初めて，問題点に気付き，解決策を講じたというほうが事実に即している。

2. 問題点の所在

この節では，電子計算機を使った語彙調査において，「語」を記述する上で，どのような問題点があるのか，具体例で考える。

語彙調査の対象として，一まとまりの文章を取り上げる。最も短く，しかも完結した文章は，俳句や短歌であろうが，これらは，句読点など補助記号を用いないこと，清濁を書き分けないこと，歴史的仮名づかいで書かれているなどの，現代語一般の文章とは異なった，表記上の制約がある。そこで，次の文を，一まとまりの文章と見なして，考えを進めることにする。

山道を登りながら，考えた。

この文章が，語彙調査の対象として取り上げられたと仮定する。そして問題を単純にするために，サンプリングは行わずに，全数を対象として語彙調査を進めるものとする。

ここで「語彙調査」がどのようなもので，何をするのかについて，規定しておく必要があろう。これについては，国立国語研究所報告13「総合雑誌の用語・後編」の「付録Ⅰ．語彙調査の成立根拠と基本的諸概念の定義」(94ページ)が，次のように簡明に述べている。

> 「語彙調査 (word count)」と呼び慣らわされて来た調査のおもな仕事は，各「語」の使用度数，更に適切には使用率を調べることであった。すなわち
>
> (1)「語」と呼ばれその存在が予期されているものについて，
>
> (2)意味上・形式上同じと見てよいものを，それぞれにまとめ，
>
> (3)まとめられた各群が実際に使われている度数または割合を測る
>
> ことだと言える。

この稿でも，この定義に従うことにする。

さて，先ほどの「山道を登りながら，考えた」の語彙調査に着手する。まず，「語」を取り出さなければならない。そのためには，調査単位を定めて，この調査対象から，その単位を切り出さなければならない。ここで，「調査対象を単位で切り分ける」と言わないのは，調査対象がすべて「語」で構成されているとはかぎらないためである。この単位の切り方については，これまで種種の方式が出され，大勢は，文節を中心とした単位と，形態素を中心とした単位の二種に分かれてきている。単位の切り方については，まだ十分に論議しつくされておらず，問題も多く，調査の目的に照らして検討する必要があるが，

ここでは問題としない。高校教科書調査のM単位で切り，切れ目を／で示す。

／山／道／を／登り／ながら／，／考え／た／。／

ここで切ったものが，すべて「語」なのかどうか，次に検討しなければならない。「山」や「道」を語とみなすことには，単位の長さの問題，すなわち「山」「道」がそれぞれ一語であるとする立場と，「山道」が一語であって，「山」と「道」はその要素にすぎず，「語」ではないとする立場との差を除いては，問題はない。この二つの立場の違いは，どちらにしても，「山道」の部分に「語」に相当するものの存在が認められるという点では，大きな差とは言えない。ここで問題とするのは，「語」が存在していると認め得るかどうかという場合である。問題となるのは，次の諸点である。

(1) 助詞・助動詞は「語」か。

(2) 接頭語・接尾語は「語」か。

(3) 記号類は「語」か。

(4) 数字は「語」か。

(5) 句読点，カッコ等の補助記号は「語」か。

例文では，「を」「ながら」「，」（カンマ・読点）「た」「。」（ピリオド・句点）を，「語」と認めるかどうか，ということである。これらのものを，従来の語彙調査では，「語」として取り上げたか，また，取り上げない場合はどう扱ったか，国語研究所の調査を順次検討する。

なお，句読点など，表記のための補助符号が，語彙調査の単位として問題となるのは，電子計算機を使った語彙調査以降である。

(1) 助詞・助動詞は「語」か

助詞・助動詞を「語」として，調査対象に含めるのかどうか。これは語彙調査にとって，最も基本的な事がらであるが，これまで十分に論じられた形跡がない。この例で「を」「ながら」「た」を「語」とみなすかどうかは，調査結果に大きな影響を与える。調査規模が大きい場合は，全体（延べ語数）のかなりの部分を占め，しかも異なり語数が少ないため，度数順に配列し場合，その上位に集中するはずである。

国立国語研究所の数次にわたる語彙調査では，調査対象の範囲として，助詞

・助動詞を除くという方向でほぼ一貫してきたようだが，子細に検討すると，微妙に食い違っている。最初の報告「語彙調査—現代新聞用語の一例—」（資料集2〈1952〉）から見ていくことにする。資料集2「1. 調査の目標と調査の対象」（3ページ）では，

> 紙面から採集するものは，いわゆる自立語だけで，助詞・助動詞はとらないこととした。

となっている。しかし，「2. 調査単位の切りとり方」（4ページ）において，

> カードに見出し語として掲げるものは，原則として「文節」を最大の単位とする。これは「附属語」をその直前の自立語とともに1枚のカードにとるということであって，それは，まず「自立語」のみを当面の調査対象としながらも，カードの取扱い方によっては「附属語」をも検討しうるようにしたものである。

となっており，付属語も検討する姿勢も見せている。しかし，助詞・助動詞の集計表は作られず，また，各種の語彙表（無活用語・動詞・形容詞・接頭辞・接尾辞のそれぞれ，および度数100以上のものにつき以上を総合したもの，そして助数詞）にも，助詞・助動詞は含まれていない。

次の語彙調査の報告「婦人雑誌の用語—現代語の語彙調査」（報告4〈1953〉）では，調査単位として「α単位」が選ばれ，単位切り作業の規定で，「原則1次の各項に該当する箇所で，文を切る。」に続いて，

> ㈣助詞，または助動詞の後。助詞・助動詞がいくつも続いている場合は，最後のものの後。(20ページ)

とあり，さらに

> なお，採集した語の排列にあたっては，α単位から助詞および助動詞を取り去った形をもってした。(26ページ)

とあり，α単位内に，助詞・助動詞が含まれるかのようである。しかし，これに続いて，

> 助詞と助動詞とについては，以上の調査とは別に扱うこととした。それは，助詞・助動詞が一般の自立語と性質を異にするものと考えられたためである。すなわち，自立語とは別個に採集範囲を決め，用例を採集し，助

詞・助動詞だけの度数表を作成することとしたわけである。

とあって，295ページ以下に「『婦人生活』の実用記事に使われた助詞・助動詞」が掲げられている。したがって，51ページ以下の「語彙表」には助詞・助動詞は含まれていない。また，315ページ以下の「8この調査の反省」では，調査対象・調査単位などについて反省を行っているが，助詞・助動詞の扱い方には触れていない。ただ，調査単位として取り上げたα単位に疑問を持ったが，

根本的に単位を切り替えるには及ばず，α単位のまま（もちろん助詞・助動詞ははずしてあるが），この報告書に掲げられることになった。

と述べられている。これらのことから，婦人雑誌の用語調査では，助詞・助動詞を，α単位の一部であると認めはしたが，語彙表では，一般の自立語とは別のものとして除外していると言える。

この，助詞・助動詞を除外するという姿勢は，次の総合雑誌調査では非常に明瞭である。ここでは調査単位としてβ単位が採用されるが，そこで助詞・助動詞は除外される。「総合雑誌の用語・後編」（報告13〈1958〉）10ページ「2・3調査単位の句切り方」において，以下のように述べられている。

【助詞・助動詞に対する処置】この調査では，いわゆる助詞・助動詞の大部分を調査対象から除外した。従って，助詞・助動詞に対しては，語源単位としては考えないし，また以下に述べる「β単位の規定」も適用しない。ただ，一般に助詞・助動詞と言われるものの中で，特に調査単位として取り上げた語例があって，それらは，後に〔別表〕として示す「一β単位とする付属要素」の中に掲げてある。

このように，総合雑誌調査は，助詞・助動詞を明確にβ単位から除外し，また，助詞・助動詞の集計も行っていない。

次の現代雑誌九十種調査では，調査単位としては，同じβ単位を採用したが，総合雑誌調査とは異なり，助詞・助動詞を明確にβ単位内に定義している。すなわち，「現代雑誌九十種の用語用字・第一分冊」（報告21〈1962〉）6ページで，

今回の調査単位には，総合雑誌の調査の時と同様に「β単位」を採用した。今回は，前回と違って，助詞・助動詞も含めて調べた。

と述べ，助詞・助動詞を「最小単位」のひとつとして分類し，作業規則において，

3　付属要素，符号，助詞・助動詞は，一最小単位を1βとする。

と規定している。そして，「助詞・助動詞以外につき，90誌全体にわたる五十音順語彙表」と並んで，「助詞・助動詞につき，90誌全体にわたる五十音順語彙表」が掲げられている。

次の新聞の語彙調査では，調査単位として「長単位」と「短単位」が採用された。「短単位」はβ単位をかなり引き継いでいるため，助詞・助動詞の位置づけは，雑誌九十種調査とほぼ同様である。ただし，結果の示し方は，大いに異なる。全体の三分の一（一紙一年分）のデータの結果を示した，「電子計算機による新聞の語彙調査」（報告37〈1970〉）・「同（Ⅱ）」（報告38〈1971〉）・「同（Ⅲ）」（報告42〈1972〉）には，助詞・助動詞を含めた全体の短単位表と，品詞別度数順短単位表の中の「助動詞の表」「助詞の表」のかたちで示されている。全体から助詞・助動詞を除いた短単位表はないが，全短単位表に「部分」の結果が併記され，それは「全体」から，固有名詞・助詞・助動詞・算用数字・記号類をのぞいた集計結果である。なぜ助動詞と助詞とを別に集計したか，なぜ「全体」から固有名詞・算用数字・記号類とともに助詞・助動詞をのぞいたか，説明はなされていない。当事者の一人として，説明は可能である。すなわち，助詞・助動詞を合わせた結果は，この報告から容易に出すことができること，固有名詞は新聞という資料の性格から非常に多く，他資料と比較する上でさしつかえること，算用数字・記号類は「語」と認めるのに難があること，などである。しかし，本論においては，あまり適切な理由とはみなせない。

以上のように，これまでの語彙調査では，助詞・助動詞を「語」として扱うかいなかについて，全体の語彙の中では除外することでほぼ一致しているものの，付属語として取り上げるもの（婦人雑誌・雑誌九十種・新聞）と，そうでないものとがある。これは，語彙調査の対象たる「語彙」に関する論議が不十分であるか，十分論議されたがその結果が徹底しなかったためと考えられる。助詞・助動詞について，土屋は付属語として取り上げるべきだと考える。すなわち，語彙調査の結果は，自立語だけの集計，付属語だけの集計，および両者

を結びつけた集計の三種を必要とすると考える。しかし，本論は対象とする「語」をいかに記述するかが中心であるので，助詞・助動詞も把握されるべき「語」であることを述べるにとどめる。

(2) 接頭語・接尾語は「語」か

接頭語と接尾語を，語彙調査の中で，「語」として把握すべきかどうかについて考える。接頭語・接尾語は，必ず自立語の一部分であるから，これまでの語彙調査では，いずれも「語」の中に含まれている点では問題はない。調査単位の長さにより，一調査単位として独立した見出し語が与えられるかいなかの違いがあるだけである。たとえば，昭和24年の新聞語彙調査では，接頭辞・接尾辞・助数詞として，それぞれの語彙表が掲げられたのに対して，婦人雑誌の語彙調査では，α単位を採用したため接頭語・接尾語はその中に含まれてしまって見ることができない。しかし「それらが実際にどのような広さで，また，どのような多さで用いられるかをたしかめることは，語彙調査の上で，一つの重要な課題になる」（「報告4」25ページ）という見地から，特に重要と思われるものを取り出し，別に整理・集計して，別表として掲げてある。また，総合雑誌・雑誌九十種調査のβ単位，新聞調査の短単位では，いずれもひとつの調査単位として，語彙表に掲げられている。

ここにおける問題点は，次の二点である。

(a)接頭語・接尾語の認め方にはばらつきがあること。

(b)接頭語・接尾語を調査単位として取り出した場合，他の自立語と併列させてよいかどうか問題が残ること。

(a)は，たとえば，「れる」「られる」「せる」「させる」は助動詞とみなすか接尾語とみなすかで，語彙調査の単位となったり，ならなかったりする。これらの言葉は，婦人雑誌調査では助動詞として扱われ，「助動詞用法別度数表」に取り上げられ，総合雑誌調査では一語源単位と認められず，結合した形をβ単位としたため，語彙表中には有るが発見は容易でなく，雑誌九十種調査ではβ単位として語彙表に見出し語を立てて掲げられている。これらはかなり使用頻度の高いものなので，語彙表に含めるかどうかによって上位の順位および累

積の使用率に大きな影響を及ぼすことになる。語の認め方について言うなら，「である」「でない」の「ある」「いる」を用言とする婦人雑誌と，助動詞とする総合雑誌・雑誌九十種調査とでは，結果に大きな差違を生ずるし，しかも総合雑誌調査は助動詞の集計結果を掲げていないため，これらの「ある」「いる」の使用状態がつかめない。これに比べれば，(a)は小さな問題と言うこともできようが，調査の実施にあたって，まず態度を明確にしておかなければならない事がらであろう。

(b)は，形式的には，語彙表において，自立語にまじって，接頭語接尾語を掲げてよいかということだが，内容的には，単語と単語の一成分である接頭語・接尾語を，語彙を記述する単位として，同等に扱ってよいかということだと思う。これは，語彙調査の単位として，単語より小さな「形態素」的なもの(「婦人雑誌」316ページで「語源的単位」，「総合雑誌」後編10ページで「語源単位」と呼ぶようなもの）を考える場合は，さしつかえないと考える。当初，土屋は単語と接頭語・接尾語を対等に扱い助詞・助動詞を除いて「語彙」を把えるのに抵抗を感じたが，現在では，助詞・助動詞を付属語として別に把えるならば，単語も接頭語・接尾語も，自立語の構成要素として，語彙量を測定する単位として同等に扱ってよいと考える。

(3) 記号類は「語」か

語彙調査の対象中に現れた記号類は，「語」として調査単位に取り入れるべきか。記号としては，A・B・C，イ・ロ・ハなど文字を使ったもの「秘」を○で囲んだ文字と図形の複合したもの，○・×のように図形のみのものなどがある。これらの記号類の扱いを作業規則で明確にしたのは，雑誌九十種調査からであり，「報告21」11ページには，次のように述べられている。

3　付属要素，符号，助詞・助動詞は，一最小単位を1βとする。

(略)

例：〔符号〕｜イ｜図｜　｜甲｜表｜　｜×｜町｜　｜◯◯(まるまる)｜社｜

｜Na｜Cl｜〔化学記号〕　｜H｜$_2$｜O｜

しかし，実際には，どの語彙調査でも記号類は調査対象の中に出ており，婦人

雑誌調査・総合雑誌調査でも，五十音順語彙表の末尾に，掲げられている。これらから，記号類も，従来「語」として調査単位に入れられていると見ることができる。

しかし，記号類を「語」とするかいなかは，十分に検討されねばならない。「語」であるための必要条件は，語形を有することと意味を有することだと考える。記号類（(5)に述べる表記上の補助符号を除く）は，必ず何らかの意義を有するが，語形はやや不安定である。○は「マル」であろうが，×は「バツ」だけでなく「ペケ」という語形であることもある。しかし「掛ける」とは別の記号であることははっきりしている。したがって○も×も「語」としての必要条件を有している。○の中に「秘」を入れた「マルヒ」，「公」を入れた「マルコー」も同様である。一方，新しい外来語，たとえば「Vs.」や「cf.」「No.」などは，まだ確かな語形を持っておらず，「＊」に「アスタリスク」の語形を与える人はまれである。「∞」（無限大）や「≒」（イコール）も同様である。しかし，これらも何らかの語形を与えて「語」として扱うべきだと考える。ただ，安定した意味・語形の見出せないもの，極端な例では，その文章中にのみ使われた記号という場合もあり，「語」には含めるものの，語形は与えないということはあり得よう。

4) 数字は「語」か

数字によって表されるものは「語」である。調査単位の定め方，語形の認め方がむずかしいものの，雑誌九十種調査までのカード方式の語彙調査では，語形を与えられ，語彙表に並べられた。たとえば，婦人雑誌調査では「キュウジッセンチ（90 cm）」「ゴジュウモンメ（五〇匁）」などの見出し語が存在し，総合雑誌調査では「報告13」の「β 単位の規定」42・2で，「532438000円」を，

5 | 3 | 2 | 4 | 3 | 8 | 0 | 0 | 0 | 円
＊　　＊＊

と切り，＊のところに「億」，＊＊のところに「万」を補い，「$\frac{1}{5}$＊」は，＊のところに「分の」を補い，それらをβ単位として採集している。雑誌九十種調査も総合雑誌調査と同様で，たとえば，語彙表「テン」付近には「テンイチ

(.1)」「テンレイ(.0)」「テンレイレイ(.00)」などの「語」がならんでいる。

数字が「語」と認められなくなったのは電子計算機による新聞の語彙調査のときで，算用数字が固有名詞・助詞・助動詞・記号類とともに「部分」を集計する際に除外された。「部分」というのは，「常識的な意味の単語（自立語）に近い」(「報告37」28ページ「『全体』と『部分』の定義」) とされるが，算用数字は，部分に入らず仮名による語形も与えられず，漢数字とも合わされず，「語」と認められていないと言える。算用数字をこのように扱ったことについ，説明はなされていないが，思うに，新聞には経済欄をはじめ数表が多く，これらに含まれる数字を「語」と認めがたかったものであろう。算用数字を「語」から除外したため，「一点」の「一」は「語」として取り上げられ，「１点」の「１」は除外されるという不合理も生じた。

算用数字を「語」として扱わなかったもう一つの理由として，語形を示すために，「億」や「分の」を補うことに抵抗があったことが挙げられる。後述するように，電子計算機を使った語彙調査では，文章を文字・記号の列として入力するため，補うことによって，表記の実態の集計が困難になってくるからである。

(5) 句読点・カッコ等の補助記号は「語」か

句読点をはじめとする各種の表記のための補助記号は，「語」とは認めがたい。数字と同じく，カード作成による手作業の語彙調査では，一切「語」としては扱われなかった。ただ，繰り返しの符号「〃」が語として「〇」などと一緒に扱われているが，これは，相当する言葉を当てるのが困難なためであろう。これも語を当てるべきものと考える。

句読点等が「語」として，調査単位に入れられるのは，電子計算機による新聞の語彙調査からである。「報告37」の「長単位の区切り方」（13ページ）から引用する。

2. 記号および記号連続は１単位とする。

2.1. 文字・数字・スペース以外，すなわち句読点・くぎり点・引用符・かっこ　ダッシュ・リーダー・疑問符・感嘆符・数字記号・音楽

記号などは，すべて記号類と認める。

2.2. つぎの記号類は無視する。

○数字連続の中に現われる小数点・位取りカンマ。

○よみがなの前後にあるカッコ・ダッシュなど。

○単に名詞を連結するための記号，たとえば姓と名を結ぶ記号など。

ここでは，調査単位と認められる記号類と無視される記号類とがある。無視されるものは，表記上の補助符号ばかりであり，これが「語」ではないことは明らかである。調査単位とされるもののうち，数字記号・音楽記号は，(3)・(4)で考えてきたように，「語」であるが，その他は表記上の補助符号であり，「語」を形成してはいない。もしこれらを調査単位とするなら，段落の先頭を示す一字下げ（スペース）やアンダーライン，ゴチック・イタリック書体等も取り上げねばならないはずである。

ではなぜ表記の補助符号の一部が，調査単位とされたのか。それは，次に述べるように，電子計算機を使った語彙調査の入力データの形式の制約によるためである。

3. 入力データの形式と「語」

前節では，「語」と認めるものの範囲について問題点を検討した。そして，電子計算機を使う語彙調査から，目立って「語」の範囲が広がっていることを確認した。この原因は，電子計算機を使った語彙調査における言語データの扱い方にあると考える。そこで，最初の例文「山道を登りながら，考えた。」に戻って，考えてみる。

この文中の句読点は「語」か。これまで検討してきたように，従来の語彙調査では句読点は「語」とはみなされず，また，「語」は語形と意味を有するものであるという考えからも，語形を有していない句読点を「語」とみとめることはできない。それでは，句読点をはずして入力してしまってよいか。それはもとのデータの形式を再現できなくしてしまうため，好ましくない。語彙調査の目的は語彙の量的な把握であるが，日本語の場合は，それに伴って表記の実態の量的把握も目的とされ，そのためには，表記形態がそのまま各語に添えら

れていることが要求される。したがって数字の位取りのカンマや小数点はともかく，その他の表記のための記号類も，調査単位に入れることになる。

句読点などの補助記号の扱い方だけでなく，文全体も，形を崩さずに入力することが要求される。その結果，さきの文は調査単位の切れ目を示す記号を挿入されただけで，入力データとしてインプットされる。このデータを調査単位の切れ目で切り，同じ形態のものを集めて整理すると，何が得られるか。これは同形態の文字および記号連続の集計である。「語」を集めた語彙表には程遠いものである。電子計算機による新聞の語彙調査の長単位表が，この文字記号連続の集計表の一例である。これは文字列表と呼ぶことができる。これはこれで文字列の集計として意義があるが，「語」の実態はつかむことはできない。

語彙調査をすすめるためには，「語」に語形を与えなければならない。さきのデータで，「語」ではない句読点には（×）を付加し，語形を片仮名で表記し，出現した文字列を語形のあとに（ ）を付して記入すると，次のようになる。

／ヤマ（山）／ミチ（道）／オ（を）／ノボリ（登り）／ナガラ（ながら）／，（×）／カンガエ（考え）／タ（た）／。（×）／

出現した文字列をカッコ中に移したため，もとの文の形態とは変ったものの，復元は可能である。このような形式でデータを作成し，機械処理を施すと，何が得られるか。これは語形の集計結果である。

さらに「語」に近付けるために，いわゆる同語異語判別作業を行って，語形の変化しているものを一つの語形のもとに集める，異形同語の統合作業と，同語形のもので意味の異なる，同形異語の分離作業を行わなければならない。同形異語の分離作業とは，例えば「ある」を連体詞「或る」と動詞「有る」に分離する類である。異形同語の統合は，例えば活用語にすべて終止形を与え，出現語形を（ ）を付してその中に移すなどの手を加えることである。この作業に先立って，「語」の単位の幅を規定することが必要である。総合雑誌調査・雑誌九十種調査では「集計単位」としてこの規定が存するが，他の語彙調査では明確でない。この操作の結果，例文は次のようになる。

／ヤマ（山）／ミチ（道）／オ（を）／ノボル（ノボリ）（登り）／ナガラ

（ながら）／，（×）／カンガエル（カンガエ）（考え）／タ（た）／。（×）／

この例文では，これで同形の別語が存在しないが，もし存在する場合は，それぞれを分離するために，意味の情報を添えなければならない。高校教科書調査の場合は，主として，出現形の先頭文字を意味情報として採用する方式をとったが，これは作業が簡単なために選んだ方式であり，本来は相紛れることのない意味情報を付加すべきであると考える。（この意味情報の記述形式に関しては別に考えたい）

上記の操作の結果，「語」を取り出すことはできたが，自立語と付属語が混在しているため，これを分離して語彙表を作るため，付属語のみに「助」を添えることにする。この結果，例文は次のようになる。

／ヤマ（山）／ミチ（道）／オ（を）〈助〉／ノボル（ノボリ）（登り）／ナガラ（ながら）〈助〉／，（×）／カンガエル（カンガエ）（考え）／タ（た）〈助〉／。（×）／

この操作を経たデータを機械処理することによって，自立語と付属語の語彙表が得られ，語彙調査は，当初の目的を達することができる。

このようにして，電子計算機を使った語彙調査の初期の段階で採られた「文字列」の集計を改め，情報を付加することにより，「語」の連続，つまり「語列」として把え，集計することを考えるに至った。付加する情報の内容の吟味，および付加処理システムの検討は今後の課題である。なお，この考えは，高校教科書調査の途中で生まれたもので，調査全体の流れとは必ずしも一致しない。

付記　本稿は，昭和53年9月計量国語学会第二十二回大会において「語彙調査における語のとらえ方について」と題して発表したものに，その後の考えを加えて，まとめたものである。