

意味分野別構造分析法^(注)

—— 語彙研究法としての ——

田 島 毓 堂

0. はじめに

総体としての語彙を分析するには、語彙の持つ二つの面、数量的側面と意味的側面を同時に満足する方法が必要である。それに堪えるのが、阪倉篤義氏が「万葉語彙の構造」（『万葉』34 1960）で試みられた方法である。私はこの方法を「意味分野別構造分析法」と名付けて、諸種の語彙分析に応用している。

本稿においては、その方法の重要なことに鑑みて、その対象語彙についての考え方、語の単位のあり方、意味分野の設定法、意味コードの付け方等を中心に、語彙研究の一層の進展のために、意味分野別構造分析法ということについてきちんと定義しておこうと思う。

語彙研究にその要素たる個々の語を研究する部門と語の集合として、文字どおりの語彙を対象とする部門のあることについて、もはや贅言の要はなからう（「語彙論の課題—集团的規範と個別的実現—」『名古屋大学国語国文学』71 1992）。また、言語研究の一般として、その個々の実現としての側面とそれを成り立たせる集団規範としての側面の両者があり、その研究が混同されてはならないこともいうまでなからう。

本稿では、語彙を総体として捉える最有力な方法であると考えられる意味的側面からの分析法を提唱した阪倉篤義「万葉語彙の構造—（その一）名詞について—」（『万葉』34 1960）において実践された語彙分析法を取り上げ、その意義を顕彰するとともに、さらなる語彙研究の進展のためにその限界を指摘し、意味分野別構造分析法の具体的内容にわたってその内容を述べようと思う。

なお、それに先立ち若干の準備として、語彙分析として実践されている、「意味分野別構造」以外の諸観点からの実情を述べてその意義と限界とを指摘しておくのも無意味ではあるまい。

1. 語彙の性格の両面

語彙は「語の集合」であることはいうまでもないと思うが、この元である一々の「語」はそれぞれ意味をもってあいより、あい集まっている。従って、この「語の集合」としての語彙には、その性格として、一には数量的側面、また、一には意味的側面があることが指摘できる。まずその数量的性格としては、その使用頻度・使用率からみた語彙の特性、品詞別の語彙構造のあり方、語構成別の語彙構造からみた語彙の特性、語種別の語彙構造からみた場合の特性、さらに、語長

別の語彙構造の特性などが指摘される。これらについての語彙調査の結果の知見は、それぞれ語彙の一般的特性を示してはいるが、他の語彙との比較による共通度調査によって得られる知見などと共に、その数値的特徴は個別語彙（「語彙論の課題—集团的規範と個別的実現—」参照）（＝テキスト語彙）を明確に特徴づけることにはほど遠い。また、異なり語数と延べ語数という数値、及びその比である一語平均使用回数がいかなる意味を持つかも、少なくとも、従来十分に明らかにされていない。仮にその数値の持つ意味が知られても、個別語彙を特徴づけるにはほど遠い。微細な数値の違いが、何を有意に示すのかを明確に意味付けることが困難だからである。仮にやはり有意に差のあるという数値が示されたとしても、そのもつ意味は限定的であろう。すなわち、語彙の一特徴である数量的側面からのアプローチだけでは必ずしも、個別語彙の特徴は明らかにならないのである。

もう一方の意味的側面からの接近が不可欠である。ただ、それも個々の語の意味の記述だけでは、ついに、語彙の分析にはつながらない。意味的側面を一方の数量的性格と結び付ける必要がある。ところが、意味は数量化に最もなじみにくい。それを数量的に扱おうとしたのが、阪倉氏の方法である。それは、意味をカテゴリー化した上で、そこに語を当てはめ、その個々の語にコードを与え、コード毎に（阪倉氏は名詞語彙について、5類23種にまとまとめてグループ化する）集計して語彙の意味分野別の構造を知り、それを語彙分析の方法とするものであり、私はこれを「意味分野別構造分析法」と称するのである。この用語については、若干不明確だと指摘があったが（湯浅茂雄「語彙（理論・現代）」（『国語学』177 平成4年・5年における国語学界の展望 1994）、注に述べたとおりである。

阪倉氏の論文についての詳細な検討と、改善案については別稿「語彙研究法としての『意味分野別構造分析法』のために一阪倉論文の顕彰と批判」（『国語と国文学』に掲載予定）に記したので、本稿では、意味分野別構造分析法の実際とコード付け等について述べる。

2. 意味分野別構造分析法のあらまし

今、簡単に意味分野別構造分析法について、「意味をカテゴリー化した上で、そこに語を当てはめ、その個々の語にコードを与え、それによって語彙を分析する方法」だと述べたが、もう少し説明しよう。

①「意味をカテゴリー化」したものがいわゆるシソーラスの各項目であるといえよう。その項目を全部集めれば、「意味の一覧表」になっている必要がある。「概念の一覧表」といっても同じ事である。全ての意味をもれなく掲載する必要があるが、それには限度もあろう。つまり、こういう「意味の一覧表」も或る特定の言語によって作られるはずであるが、そうすれば、その言語においては認識されないような意味領域は抜け落ちる可能性がある。できるだけ他言語を参照してその欠落を防ぐ必要があろうが、あらゆる分類がそうであるように、「その他」という項目は避けられまい。幸い、日本語には、ロジェのシソーラスを参考に作られた『分類語彙表』とい

うものがあり、これは細部に改善すべき点はあるとはいふものの、よくできた「意味の一覧表」になっている。阪倉氏によれば、『分類語彙表』の前身である林大氏の分類案を「日本語の意義分類のうち、科学的なものとしてはほとんど唯一の先蹤」と称している。

「意味をカテゴリー化」したものとして、『分類語彙表』を使うことは単に一つの基準、共通の基準という以上のものがあると思う。もっとも、意味分野別構造分析法のためには、個々の語にそれによってコードを付けること、言い換えれば、その基準になる『分類語彙表』にあてはめて、該当するコードを付ける必要があるが、『分類語彙表』は意味分野別構造分析法のために作られたものではなく、本来、類語辞典の性格を持つもので、その方面での使用が大きな意味を持っていたものである。

それ故、意味分野別構造分析のためには、『分類語彙表』を利用しながら、種々の工夫がいることは当然のことであり、ひいては、それが、『分類語彙表』の改善にも繋がっていくものと思う。意味分野をどのように設定するかは、種々の方法があり得る。『分類語彙表』の意味分野の配列はその一つを示したもので、言葉を体・用・相・その他に四大別し、さらにその中を、1. 抽象的關係、2. 人間活動の主体、3. 人間活動、4. 生産物および用具物品、5. 自然および自然現象に分類し、それぞれを細分類してある。四大分類は文法的観点による分類であり、鈴木朗の「言語四種論」にいう四種ともほぼ同じであり、多くの言語に於いて通用する普遍的な大分類である。次の五中分類は、ロジェ以来のシソーラスに範を取って分類したもので、それが、四大分類と連係しているところが、『分類語彙表』の大きな特徴である。ただ、細部においては、若干の食い違いも出ているが、『分類語彙表』の類語辞典的性格と目的からいえば、特に問題になることではない。しかし、意味分野別構造分析法にとっては、細部まで整合性がとれる方が望ましい。若干の修正や項目の増設なども必要になると思う。『分類語彙表』では、主として、その項目所属語の多寡により、小数点第4位が設けられているところと第3位・第2位までのところがあり、若干、体・用・相の間の連係の取りにくいところがある。必要に応じて、このあたりに工夫を凝らしていく必要がある。また、『分類語彙表』を基準として使う以上、言っても仕方ないことかも知れないが、本来は、この四大分類と、五中分類は逆であってもよかったかも知れない。もちろん、四分類の間に連係が取れているから、あまり問題はない。

なお、『分類語彙表』はこの意味分野別構造分析法にとって有用な基準であることは上述の通りであるが、現代日本語の共通語をもとにしたものでありこれが唯一絶対のものでないことは言うまでもない。将来、もっと意味の一覧表として、意味分野別構造分析法のためにも使いやすいものができることも当然考えられるものであることを付け加えておく。

②つぎに、「そこに語を当てはめ、その個々の語にコードを与え」ということは、対象とする語彙の個々の語について、それぞれが、『分類語彙表』の項目のうちのどこに属するかを見極めて、そこに配当することを意味する。『分類語彙表』におなじ言葉がある場合はほとんど問題ない。それでも、対象とする語彙の中での意味と、『分類語彙表』での意味とが食い違ってい

る場合には、『分類語彙表』のコードが直ちにはその語のコードにならない、その語にはふさわしくない場合もあり得る。それも問題になるが、そのことよりも、『分類語彙表』が現代語を基礎にしており、しかも、そこに収載されているのが三万三千語弱、各語には一コードを与える、ということから、多くの複合語はすぐにはコードが求められないし、たとい複合語があったとしても、そのコードは語の一部しか表さないという複合語の取り扱い上の大きな問題がある。複合語は多くの場合、その語の構成要素のうち後部要素によってコードを与えられている。その結果、その語の前部要素、つまり、後部要素に対する修飾部分は無視されることになる。このことは、一語に一コードを付けるということならば仕方ないことで、『分類語彙表』による限りは何ともならない。『分類語彙表』では、一語の意味を過不足無く、如実に表すほどには細かく分類されてはいない。これをこのまま放置しておいていいかどうか。このことについては、後に述べることにするが、「意味」を分析の中心に据える意味分野別構造分析法にとっては、実際にある部分が無視してしまうことは重大な欠陥を生じることになるのではあるまいかと思われる。

③「コード毎に集計して語彙の意味分野別の構造を知り」、とにもかくにも全ての語にコードを付け終わったとして、一々のコードについて、それぞれに何語あり、同時に何回使われているかを把握する。すなわち、その語彙に、どういう意味分野の語がどれだけあるかを示す表ができる。これが、すなわち、その語彙の意味分野別構造そのものを示す。言うまでもないが、全て数字の表である。これだけでは、その語彙の意味分野別の構造が数的に示されただけで、単独にそれを見ているだけでは、何事も明らかにならない。「それによって語彙を分析する方法」といっても、具体性をもたない。結局は、何か他に対照するものが必要になる。何か標準的モデルがあって、それと比較して、どの部分がどうで、この部分がこうだということがいえればいいが、現在のところはそういうモデルはないし、よほどの蓄積を経た上でなければできないものではないだろう。さしあたりは、ジャンル別のモデルを考えるとところから始まるだろう。もちろん、完全な標準になるようなものは、完全な標準としての「語彙」というものが考えられない限り、無理である。個々別々の語彙(=作品)はあっても、「標準」といった語彙(作品)は考えにくいであろう。ただ、代表的なものということであれば、考えられないこともなかろう。そういうものができれば、それと比べて、この作品の語彙はこれこれの特徴を持つということが言えるであろうが、当面、何らかの対照語彙について同様の分析をしてそれと比べてみるということにならざるを得まい。

つまり、意味分野別の分布表をもとにして、作品の語彙ごとにその意味の「野」における分布の状況をあきらかにした上で、それを分析の手がかりにする試みである。

そして、そのそれぞれにおける語彙の意味分野の分布状況の違いは、その背景にある思考形態の相違を反映するものであると考えられるのである。

2. 1. 0 意味分野別構造分析法実行上の留意点

上に、意味分野別構造分析法について、三点に分けて、その内容について述べた。その一つについて、実際上の問題点を指摘しながら、その対処の仕方についても述べる。

2. 1. 1 分類語彙表—意味分野設定の基準として—

「意味をカテゴリー化」した、意味の一覧表として、『分類語彙表』は基準として適切なものである。もちろん基準としては、必ずしも『分類語彙表』には限らないが、既にこれを使ったいくつかの研究がある。このほかにも、例えば、大野晋・浜西正人氏編『類語新辞典』1981、『類語国語辞典』1985がある。『分類語彙表』とは全く違う原理で作られている。これに依らないのは、内容的な問題ではなく、これを使った研究がなく、成果について比較検討することができないからである。

『分類語彙表』を使って個々の語にコードを付ける際に留意すべき事が、大きくいって二点ある。その一は、『分類語彙表』が現代語共通語の調査からできていることから、これによって古典語・方言について、コード付けをするには、種々の困難があるが、これは本質的なことではない。同一の意味の現代語があれば、現在の『分類語彙表』の仕組みの上からは、それに比定してコードを付けることになる。その二として、現代語を対象とした場合にも、その同じ語が『分類語彙表』に収載されていないということが問題として起こる。主として複合語と方言についてである。『分類語彙表』の収載語数は三万三千語弱、現在増補改訂の作業が行われており、その評価版が出ている。八万五千語強に増補されている。しかし、いかに増補しても、あらゆる語に対応できぬ事は言うまでもない。コードの付け方が問題になってくる。

これとは別に、『分類語彙表』ではカテゴリー化されていないが、適切な位置に新たにカテゴリー化してコードを新設しておいた方がいいものがある。『分類語彙表』は意味分野別構造分析のために作られたものではないから当然であろう。その観点から、若干の修正とコード新設することは、そのことを明らかにした上で、する必要がある。

2. 1. 2 コードの付け方

「そこに語を当てはめ、その個々の語にコードを与え」ることには実は色々の問題がある。前項で述べた『分類語彙表』にその語がない場合の外にも、意味分野別構造分析にとっては重大な問題がある。これをきちんと処理しておく必要がある。

『分類語彙表』は一語に一コードである。一語に二つ以上のコードがあるのは、多義の場合で、それぞれの意味に対してコードが与えられているのであり、多義でない語には、いくら長い語でも、つまりいくつかの要素の組み合わせされた語でも、一コードである。むしろ、複数の要素からできている語ほど意味的には限定されてくるから、多義語ということは考えにくい。という

ことは、複合語の中のいくつかの要素の意味は無視されるということが起こる。例えば、『分類語彙表』の「1.525 川・湖」には「川・河川・河流」のほか「小川・大川・谷川…清流・濁流…池・古池・沼・沼地…」といった語がある。「小川」と「大川」、「清流」と「濁流」とが同じコードである。これを区別するコードはない。しかし、意味を分析の観点とするのに、複合語の修飾部分が全部欠落してしまつては、意味的に欠陥を生じかねない。これは重大な問題である。そのために複合語の個々の構成要素を「意味単位」として認め、その一々に、コードを与えることを提案し、実際に試してみた。その結果やはり、両者には差が出た。どちらでも同じならば、簡単な方がいいが、はっきり差が出る以上、忠実に意味を反映する方法を採るべきである（田島毓堂「語彙論的語の単位試論—意味単位と分類単位と—」『日本語論究 2 古典日本語と辞書』1992）。

他言語の語彙を比較の対象とするときには、単に「意義質」（泉井久之助「語彙の研究」『国語科学講座』昭和10）のみでなく、「文法質」も見逃せない。『分類語彙表』には若干の助詞や助動詞・接辞もあるが、原則的には対象にされていない。しかし、助詞や助動詞も名詞や動詞とは質的に違うかも知れないが、語彙の要素である。語彙という観点から見れば、言語の全ての要素が、語彙の要素でもある。しかし、それを取り上げることができるかどうかは別問題であり、できることからやっていく必要がある。声の大小や音色、話し言葉としては意味を持つ音調、表現上は大切な文字の大きさや位置、語順による文法的役割、などは現在のところコード化できない。大ざっぱに言えば、文字化できることが、コード化のめどになる。この場合、「意味コード」という名称に違和感が生じる。そこで、「語彙の要素のコード」ということで、「語素コード」と称し変えることにした。従来、「意味コード」と称していたものも一括「語素コード」に移行する。その概略は、田島毓堂「比較語彙論のために—調査単位とコード付け—」（開発・文化叢書21『比較語彙研究の試み』名古屋大学大学院国際開発研究科 1997）、田島毓堂・広瀬英史「語素コードに関する提案—比較語彙論のために（その2）—」（『語彙研究の可能性』語彙研究法報告2 名古屋大学大学院文学研究科 1997）に述べた。

この複合語の扱いには諸先学も難儀をしている。その意味の欠落を防ぐために、種々の工夫をしている。阪倉氏は、できるだけ複合語の認定を厳密にすること、つまり、意味の欠落をできるだけ少なくしようとされた。阪倉氏に次いでこの方法を発展させられた浅見徹氏も、やはりできる限り複合語を認めない方針を取られた。要素に分割してしまえるものは、できる限り分割されたが、分割すると、その要素の意味の合計と、もとの意味とが合わないような場合が出てくる。「朝北」「はなむけ」などを「朝」と「北」、「鼻」と「向け」にしてしまつては、「朝北」の持つ「風」の意味、「はなむけ」の「餞」の意味は出てこない。こういうものや、他の若干のものを分割せずに残された（浅見徹「古代の語彙Ⅱ」『講座国語史 3 語彙史』1971）。私も、上記の「意味単位」を提案した際にはこの浅見氏の方針を踏襲したが、なお、不徹底の感を免れなかった。考えてみれば、複合語は橋本進吉氏が「国語法要説」（国語科学講座 昭和9 『橋本進吉博士著作集 2 国語法研究』所収）で「語の構成要素としての語及び接辞と語根」として、

「語（単語）は、更にそれよりも小さい、意味を有する言語単位に分解する事が出来るかどうかといふに、「やま」「かは」「おもふ」「ゆく」「あゝ」などの語は、これ以上分解出来ない。かやうなものは甚多いのであるが、語の中には、また分解出来るものもある。」「それぞれ意味をもつてゐる部分に分解出来る」が「既に合して一語となつた以上は、もとの語はその独立を失ひ、新たな語の部分を成すに過ぎないものである。」その証拠として、「四、意味に於いても、複合語の意味は、もとの語の意味が加はつただけでなく、それが結合して新たな意味が加はつて全体として一つの意味を表はす。「あまがさ」は「雨」と「傘」だけでなく「雨のふる時用ゐる傘」の義であり、「ほんばこ」は「本を入れる箱」であり、「はげあたま」は「禿げた頭」であり、「あかおに」は「赤い鬼」である。」とされているように、もとの語の意味が加わっただけでなくそれが結合して新たな意味が加わった全体として、一つの意味を表すことを要件にしている。つまり、要素の合計と複合語の意味との乖離は程度の差と考えられる。それゆえ、比較語彙論的考察の試案として発表した「源氏物語と絵巻詞書の語彙」（『日本語論究 4 言語の変容』1995）では徹底して分割して考えた。ただ、こうするとその複合語が成立した基盤を知ることにはできるが、その語そのものの意味が無視されることになる。これを避けるためには、再び、もとの一語にコードを付けて観察する必要がある。この両者を併せて複眼的に見ることが必要である。ただ、そういっても、もともと、複合語に、その意味を過不足無く如実に表すことのできるコードを付けることは至難であつた。そのために要素に分割したのであつた。しかし、両コードから眺めるといふことから、一つだけのコードに頼っていたのよりはよほど安心できる。これでも十分であるとは言えないが、現在考えられる方法としては、最善であるといえなくとも、これ以外のやり方よりはよいといえる。この、一語に付けるコードを「単語コード」と称しておく。

なお、語素コードによって、文法質までを対象にする場合、「意味分野別構造分析法」という用語そのものも実態と距離を持つことになる。どうしてもこの用語では具合が悪いときには、「拡大意味分野別構造分析法」と称することとする。ただし、こう言わなければならない事はないように思う。「意味分野別構造分析法」で大体は通用する。両者を対比する場合に必要なくらいである。

このコード付けはなかなか厄介な仕事であるが、きちんとした基盤さえできればある程度まで、電算化できる部分である。その面での基盤整備が、この意味分野別構造分析法の発展の上には、ひいては語彙研究発展のために大切なことと思う。

2. 1. 3 分析のための手法

「コード毎に集計して語彙の意味分野別の構造を知つ」たあと、それが表す意義を明確にすることが、意味分野別構造分析法の最も大切なことである。

これは、数字の表として表現される。どのコードに属する語が何語あって、さらに、何回使われているか、そして、その率はどうかといったことが、「語彙の意味分野別構造」として示さ

れるのである。それは数字の羅列である。我々はその数字の持つ意味合いを考えなければならない。しかし、これだけ見ても何も分からない。作品の語彙の意味分野別構造がどうなっているか示すデータが他にほとんどない状態だからである。阪倉氏の場合は万葉集の名詞語彙の意味分野別構造を、5類23の項目に分けて示した後、その意味合いを考えるべく古今集の名詞語彙と同様に調査してその意味分野別構造を同じように示し、それとの比較の上で、万葉集の名詞語彙の意味的分野別構造の特徴を描き出している。それより先、樺島忠夫氏は品詞的語彙構造と文体が関係することを現代語について明らかにしている（樺島忠夫「現代文における品詞の比率とその増減の要因について」『国語学』18 1954 「類別した品詞の比率に見られる規則性」『国語国文』24-6 1955）。また、大野晋氏が、品詞別の語彙構造が作品のジャンルを反映するという事を明らかにしていた（大野晋「基本語彙に関する二三の研究—日本の古典文学作品に於ける—」『国語学』24 1956）。それに依れば、歌集としては万葉集しか取り上げられていないが、歌集一般として古今集も、そういう観点からは相違のないことになるはずのものである。それが、阪倉氏の考察に依れば、意味分野別には、はっきり別の構造を持っていたことが知られたのである。

ただ、出来上がった数字の表をみても、簡単にその意味合いは知られない。比較する項目が、数項目ならば直感的にでもその多い少ないは分かる。それでも、それが意味のある差かどうかについてはよほどの開きがなければ断言しにくい。例えば、5%と8%、35%と38%はいずれも同じ3%の差であるが、その持つ意味合いは素朴に見ても違うだろうが、それをはっきり言いきるのには簡単ではない。もっと小さな数値の差についてはそれを違っていると見るか、そう見ないか水掛け論にも成りかねない。それよりも、細かいグループで分析すれば、コードの数は数百にもなるから、それを比較しても、よほど特徴のあるもの以外はどこに差があるかさえ分からないだろう。そのために、既に阪倉氏も用いられた統計的手法の助けを借りることになる。カイ二乗検定という方法である。カイ二乗の値を計算することは、コンピュータに依らぬとすれば、大変なことで、数百項目の計算などはあきらめることになるが、コンピュータにとっては特に問題ではない。それによって、どれだけの危険率、言い換えれば、どれだけの確率で、有意差があるといえるかが科学的・数値的に示される。色々の比較の方法があるが、要点は以上に尽きる。そして、有意差の指摘できる項目について、個々の語の検討まで進み、その持つ意味合いを考察する。これは、対照する語彙との比較においてという限定付きの特徴である。それは、比較対照する語彙を増やすことにより、段々普遍性を得ていくことになる。意味分野別構造の標準モデルのようなものも、いずれは考えられるであろう。その前段階として、作品のジャンル別の標準モデルといったものも考えられるだろうが、言うべくして至難の業である。不可能だとは言えないまでも。

3. 1 対象語彙—集团的規範語彙と個別語彙—

対象語彙として、個々の作品・言語資料の語彙をとる場合と、一言語体系の語彙を考える場

合とがある。この両者は全く異なるから、はっきり区別しなければならない。両者を混同してはならない。パロール的語彙（これを「個別語彙」と称する）と、ラング的語彙（これを「集団的規範語彙」と称する）との区別である。個別語彙はその全体を把握することができる。だから、全体を対象にすることもできるし、何らかの方法で、その一部を抽出して考察することもできる。しかし、集団的規範語彙はその全貌を完全を知ることはほとんど不可能である。その体系も明らかになっていないから、一部を抽出することも、その方法を十分考えなければならない。しかし、不可能ではない。複数の言語で、同じ方法によって抽出された集団的規範語彙は、その言語間で比較することができる。

語彙調査において、地名・人名など固有名詞が除外されることがよくある。何故除外するのか、何故あってはいけないのかが明示的に述べられないことが多いが、確かに集団的規範語彙においてはあまり意味を持つまい。しかし、個別語彙においては事情が違う。固有名詞も重要な意味を持つ。確かに、人名は「藤原」「源」とあっても、決して、「藤の原」でもないし「水の元」の意でもない。人に付けられた記号であり、いわゆるラベルである。この意味で、一般の普通名詞と異なることは確かである。しかし、普通名詞にしても事物に対するラベルであることは同じであるが、固有名詞との違いは、その語がただ一つの事物に対応するのか、其れによって複数の事物が指し示されるのかといったことが、両者を分けるポイントになろう。固有名詞は普通名詞とは、確かに違っている。それだからといって除外していい理由にはならない。それらは、人を表す語、地名を示す語として位置づければ、特段の問題はない。地名についても同様であり、固有地名は、『分類語彙表』では項目化もされている。人名・地名がいかに形成されているか、その関係はどうかということを考える材料にもなる。固有名詞は、個別語彙において重要な要素になる。さらに、これは、他言語と比べるときには興味あるテーマになる。固有名詞を、語彙調査において多くの場合除外するのは、恐らく、個別語彙としての語彙ではなく、集団的規範語彙としての語彙について、知見を得ようとする意図に出たものであろうと推考するのであるがいかがであろうか。個別語彙を対象にする場合には、固有名詞は除外されてはならない。

集団的規範語彙と個別語彙とが混同された例を、阪倉氏の論文の中から例示してみよう。

阪倉氏の論文は、万葉集の名詞語彙を対象にした研究であるが、その語の採取方針の中に次のような項目がある。すなわち、「三」として、訓仮名の用法によって存在が推定される語を、「その存在を推定し得る」語として採用すると述べる。例として、「胡粉（しらに）」「氈（かも）」「鶴（つる）」があげられる。その実、これらの語は、万葉集では、「しらに（白粉）」は「知らに」の意で使われているものであり、「かも（氈）」は「敷物」の意味ではなくて助詞の「かも」、「つる（鶴）」は助動詞の「つる」を表しており、実際には「しらに（白粉）」も「氈（かも）」も「鶴（つる）」も存在しないのである。これは、考えてみれば、実に重大問題である。恐らく、阪倉氏がこの論文を書いたときに、個別の実現としての語彙（＝個別語彙）と、それを成立させる基盤としての集団規範としての語彙（＝集団的規範語彙）の区別について意識して考えられなかっ

たのだと思う。この方針のもとに、実際にどれだけの語が算入されたのかは不明であるから、結果にどう関係しているかは分からないが、とにかく、無いものをあるとしていることになるのである。集団的規範語彙と個別語彙とをはっきり区別しなければいけないことは了解されるであろう。

いわゆる多義語の問題も、この二つの区別と関わる。よくコード付けに関して、「文脈中の意味をどうするのか」「多義語にはどのようにコードを付けるのか」という御質問を頂いた。従来、この問題は、はっきり答えられていなかった。深刻で、重大な問題であった。このことを放置した上で、いくらコード付けをして集計しても砂上の楼閣である。これはどう理解され、解決されるべきか。

実はこれも、個別語彙と集団的規範語彙をはっきり区別することによって、明確に処理できる。すなわち、個別語彙においては、個々の語は確かに文脈を有する。その意味によってコード付けする必要がある。従って、一語形であっても、別のコードを付けることは当然出てくる。「もの」は「1.100」「1.101」（抽象的關係）「1.202者」（自他）「1.400物」（物品）という4コードになる。延べ語数で考えるときは、それぞれ別々にいくつあるかも調べなければならない。これが要するに多義語であるが、個別語彙として考えるときには、それぞれ文脈中出现するときには、そこでは、その中のどれか一つとして実現しているのである。どの意味かよく分からぬということはあるかも知れぬが、とにかくどれか一つの意味で実現している。こういう意味でもあり、また、違う意味でもあるということは実際の文脈の中では特殊な場合を除いては起こらない。そういう意味で、個別語彙において文脈中に「多義語」というものはないのである。多義語は集団的規範語彙として存するのである。もっとも、こうは言っても、集団的規範語彙も個別的実現から帰納されたものである。集団的規範語彙として考えるときには、辞書が全部の意味を記述することを意図するように、その意味に対応する全部のコードを付けることになる。このように考えることによってはっきりすると思う。ただ、これに付随した問題として、「掛詞」がある。水谷静夫氏は、掛詞はその一方の意味で採用するとされる（『基本語彙と語彙調査』『国語教育のための国語講座』4 1959）。語彙計量の問題として一語を二語以上に数える不合理性を顧慮されたものであるが、そうすれば別の問題が起こる（単に一語を二回数える不合理ということだけでなく、言語の線状性に関わるという指摘もある。しかし、発話は確かに線状的であっても、その理解は話を全部聞いてから、ということもあり、後で、先に話されたことの意味付けをすることは普通にあることである。「話は終わりまで聞け」というのはこのことを示している）。掛詞は意識的に二つ以上の意味を持たせて使用されるものであり、一方を無視すれば作品の如実な理解にはならない。物体であれば、一カ所に同時に二つの物は存在し得ない（＝質礙する）が、意味には質量がないからその心配はない。あるものを無視する不合理は避けるべきであると思う。掛詞にしろ、駄洒落にしろ、個別の実現において多義的に存在しているではないかと言われるかも知れないが、これは、むしろ、普通のコミュニケーションを混乱させるところに意味を見出して楽し

むもので、特殊な場合に属する。

ただ、用語としての「多義語」は、例えば「もの」が、一つの作品内で、上掲の4つのバリエーションを持ちうるように、全体としては、一語がいくつかの意味で用いられることは当然あるものであり、こういう場合に当てはまる。ただ、はっきりしておきたいことは、同一箇所、一語が「多義」ではないという、言ってみれば至極当然のことである。

3. 2 異なり語と延べ語

語彙を分析するのに、阪倉氏以前の調査も含め、単に、異なり語数のみを問題にし、延べ語数を問題にしないものが、多かったが（後の浅見氏「古代の語彙Ⅱ」は両者を取り上げる）、一方が他方を代表できるという保証は全くない。そもそも、この両者の定義については疑問はないが、両者が何を代表するか、何を示すかということについて、はっきりした見解が示されていない（田島毓堂「異なり語と延べ語」平成7年度秋季国語学会研究発表大会はこの問題を取り上げた）。

宮島達夫氏他の『古典対照語彙表』フロッピー版によって、同表の14作品について異なりと延べを示すと表Ⅰのようになる（「語彙指標—語彙の数量的側面と語彙研究への視点—」『日本語研究と日本語教育』1992）。

表Ⅰ 語彙指標

作品名	異語数	延語数	平均	F 1	F 1 率	同延率	N	N異	V	V異	M	形異
方丈記	1148	2527	2.20	782	68.12	30.95	51.80	57.5	29.70	29.2	17.50	8.1
土左日記	984	3496	3.55	514	52.24	14.70	49.80	55.1	34.50	30.5	15.30	8.4
竹取物語	1312	5124	3.90	712	54.27	13.90	45.00	44.3	39.90	40.3	14.60	9.7
伊勢物語	1692	6931	4.09	925	54.67	13.35	49.30	54.6	36.90	31.9	13.50	8.6
更級日記	1950	7243	3.71	1105	56.67	15.26	46.80	49.1	34.40	35.5	18.60	10.8
紫式部日記	2468	8737	3.54	1408	57.05	16.12	52.50	50.7	28.20	33.9	19.10	11.6
古今集	1994	10015	5.02	1033	51.81	10.31	53.10	55.1	34.90	33.0	11.20	6.8
後撰集	1923	11955	6.21	885	46.02	7.40	50.90	53.2	36.30	34.9	12.20	7.1
徒然草	4240	17112	4.03	2464	58.11	14.40	49.50	59.1	32.50	29.4	17.80	8.1
蜻蛉日記	3598	22398	6.22	1885	52.39	8.42	40.70	47.1	38.10	38.2	20.50	10.7
大鏡	4819	29212	6.06	2495	51.77	8.54	53.10	64.0	29.60	25.8	17.00	7.2
枕草子	5246	32905	6.27	2855	54.42	8.68	44.20	53.5	32.70	34.7	23.40	8.9
万葉集	6505	50070	7.69	3246	49.90	6.48	53.00	59.7	34.30	31.4	11.40	3.8
源氏物語	11419	207792	18.19	4693	41.10	2.26	41.60	42.5	32.90	44.6	25.10	10.4

この表は、延べ語数の小さいものから順に並べてある。異なり語数（異語数）・延べ語数

(延語数)・一語平均使用回数(平均)・度数1の語数(F1)・度数1の語の異なり語数に占める率(F1率)・度数1の語の使用率(同延率)・名詞使用率(N)・名詞の異なり語数に占める率(N異)・動詞使用率(V)・動詞の異なり語数に占める率(V異)・形容詞形容動詞副詞使用率(M)・形容詞形容動詞の異なり語数に占める率(形異)を示しておいた。

異なり・延べの関係は、全体でも、品詞別でも決して平行の関係にはない。NとN異で見てみよう、多くは、N異の方がNよりも大であるが、全部が全部そうではないし、大きいといってもその大きさの度合は区々である。両者の関係は単純ではない。異なりと延べとの示すところを極く簡単にいってしまえば、異なりはその語彙の世界の広さ、つまり、素材の範囲を表し、延べはそれと共に、その世界がいかに表されているかをも示すといえよう。作品(語彙)の本質に、より深く関係するのは異なりよりは延べであるといえよう。すると、延べは異なりをも包含し得るようにみえるが、実は延べだけでも十分でないことは、個別語彙に含まれる頻度1の語が非常に多く、しかも、その使用率に占める割合はさほどでもないで、延べだけで語彙を分析すれば、その最低頻度語が無視される危険性が高い。その実それは、無視できないものをもっているのである(石井正彦「使用頻度”1”の語と文章」『国立国語研究所研究報告集17』1996 広瀬英史「使用頻度1の語彙の意味構造分析—『源氏物語(桐壺巻)』の原文と谷崎潤一郎訳を用いて—」『日本語研究センター報告5』1998)。従って、語彙を過不足なく如実に観察し、分析するには、両者の観点からする必要がある。現に、異なりと延べとでは示す内容が違ってくる(田島毓堂「語彙論的語の単位試論—意味単位と分類単位と—」『日本語論究2 古典日本語と辞書』1992のグラフ参照)。

3. 3 複合語の問題

すでに、複合語の扱いについては、2. 1. 2においても触れた。そこでは、主として意味の欠落を避けることを意図して、複合語の構成要素ごとに「意味単位」として認めてコードを与えることを中心に考えた。従来の先学の扱い方についても触れたが、それを徹底するために、すべてを要素ごとにコード化することを方針とした。ただ、複合語をどこまで認めるかは、他の語彙と比較する上で、均衡のとれたものにする上でも重要な要件である。複合語の中には、一つの事物に対応する語として、例えば「竜宮の乙姫の元結いの切りはずし」が藻の一種の名前としてあるものなどが引用される。これは、確かに一つの物の名前かも知れないが、語として一つにするかどうかは別の問題である。これに限らず、人の名前・会議体の名前・法律の名前・学校の名前等々、長大なものがある。その中にはいくつもの語を含んでいる。国立国語研究所の新聞の語彙調査の単位として「長単位」というのがあるが、これは、字数制限があった。こういう長い単位は、個別語彙では確かに特徴を表すものとして有益であるが、言語の単位として認めるには困難なものがある。時事用語辞典や用語解説辞典のためには必要な単位で、その意味では有用であるが、語彙研究の単位としてどうか、ということを考えなければならない。比較するものどうし、

同様な調査単位で行われていればいいとも言える。しかし、それはできる限り普遍性のあるものであるのが望ましい。私は、このような事物に対応した単位を「名辞単位」と称することとしているが(田島毓堂「三たび語の単位について―「運用単位」の提案―」『言語学林1995-1996』1996)、これを全く排除することは、個別語彙の特徴を表しにくくするが、かといって、無制限に認めることは、特に、漢字語について収拾のつかぬ事になる恐れがある。認めるとしても、「長単位」でしているような字数制限を伴うことが現実的であろう。ちなみに、そこでは最長十四音節に制限されている。比較する語彙どうして語数ということを問題にすることがなければ、こういうことはあまり問題にする必要がないのが、意味分野別構造分析法の長所である。つまり、すべて、意味要素の単位(意味単位)にしてコードを付けるからであるが、時には、語数そのものも問題にならぬとも限らないから、できる限り、調査単位は統一しておいた方が望ましい。

3. 4 代名詞のコード

『分類語彙表』では、代名詞には代名詞としてのコードが与えられている。代名詞は、文脈により、それが指し示すものは当然変わってくる。しかし、それに関せず代名詞としてのコードを与えるべきであると考え。これはむしろ当然のことと考えられるであろうが、阪倉氏はそうされなかった。すなわち、「いはゆるコソアドに属する語もその指示するところの実体が、場所であるか、ものであるか、人であるかによつて分類した」とある。この処置は、対象を名詞語彙としたことと関連があるのかも知れぬが、代名詞の扱いとしては是認しがたい。しかも、阪倉氏は異なり語によって考察されているから、一つの代名詞が、色々なコードを持った可能性もあるはずだがそのことについては言及されていないので、具体的処置については不明であるが、この方針は貫くことが難しいのではないと思われる。代名詞は代名詞として存在しているのであり、その指示するものに還元することは、代名詞を代名詞と見ないことになる。

4 統計的处理

統計的处理については、カイ二乗検定が阪倉氏によって有意差検定に用いられたが、判断が恣意的にならぬために、また、多数の項目の有意差検定のために欠かせぬ手法であると思う。ただ、その危険率の取り方について、何パーセントまでを有意であると考えるか、言語研究において定まった決まりはない。このカイ二乗検定によって、有意差のある項目を検出する際に、その危険率を何%にとるかということは、統計学的には一応の決まりがあり、それを参考にすべきであるが、これを語彙分析に使う場合、危険率をどのあたりまでにするかについて、十分経験を積むべきであろう。また、一方がゼロである場合がある。そういう場合の取り扱いについても、決めておく必要がある。一方がゼロであっても、他方がいくつかということ、全体の大きさが関係してくる。1対0と500対0とでは、当然持つ意味合いが違ふだろう。たとい、数学的には、共に同じだとしても。

それと共に比較する語彙の規模の問題がある。異り語数が大きくなると、使用率の小さい語を含む項目の百分率は、大きくなる。確かに、延べ語数における割合よりも、異なり語数における割合の大きい部分（項目）、すなわち、一語平均の使用回数の小さい部分ではそれが目立つ。このことは、比べる語彙の規模次第で、異なり語数と延べ語数で出現に逆転が起こることも意味する。語彙を比較して、出現に有意差が見られる場合も、異なり語の場合は、語彙の規模ということに影響されることが大きい。有意差が検出できる項目についても、なぜ有意差があるのかは子細に検討されねばならない。統計的処理は、有意差のある項目を検出するだけである。それにいかなる意味があるかを考えることが、意味分野別構造分析法にとって最も大切なことで、これが欠ければ、九仞の功を一簣に欠くことになるのである。

なお、語彙研究のためにどのくらいの量が必要かということも考えなければならない。あまりに少数の語では統計処理はなじまない。私の乏しい経験では、延べ2000語以下の語彙は統計処理の際、不安定な数値を示すことがあり、それ以上の数が必要であるように思う。

5 むすび

語彙を分析するには、それが持つ種々の側面から考えてみる必要がある。語彙の膨大性・つかみ所のなさから、一つの観点からの追究ですら容易でないということがあり、依然として、言語研究の諸側面に比して大いに遅れている。まだまだ種々の試みが成されなければならないはずであるが、本稿においては、語彙の二つの側面、つまり、数量的側面と意味的側面の両者を満足させる意味分野別構造分析法について、その考え方と方法について述べた。意味カテゴリーをさらに細分化する必要もあるかも知れない。個々の語に付けるコードについて、意味カテゴリーのコードだけではなく、さらに、語種や語構成あるいは、いわゆる位相に関するコード、語用論に関するコード等々、種々の要素をコード化することによって、個々の語の諸特性を示すものが開発されれば、意味分野別構造分析法のためのみならず、色々な面に応用できると思う。実現を期待したいが、個人的作業としては過大な課題である。

なお、この意味分野別構造分析法によって指摘される出現の有意差は「意味カテゴリー」であり、一語一語を指し示さない。つまり、かなり肌理の荒いものである。しかし、総体としての語彙研究が目指すところは究極的にはその一語一語の問題に帰するものである。ただ、実際的には、大変大まかな指摘に終わることもあり、かなりラフな面もある。さらに、精緻化する必要があるが、ともかくも、総体としての語彙に対する分析法として、この意味分野別構造分析法は活用されるべきものである。

(注) 比較語彙研究における語彙の分析法として、従来、種々の論文等で、語彙比較研究の方法として、「意味構造分析法」という言い方をしてきたが、この用語は、前々から、よく分からないと批判があっ

た。いろいろ釈明しながら使ってきたが、用語は「名詮自性」を旨とすべきであり、今後、これを「意味分野別構造分析法」(The Structural Analysis of Vocabulary with Special Reference to Semantic Categories)と改称することにした。従来の「意味構造分析法」を言い換えたもので内容的には全く等しい。今般、湯浅茂雄氏から頂いたご提案を有り難く頂戴して、今後、このように使っていくと思う。皆さんに迷惑を掛けることになると思うが、ご理解を乞う。なお、この英文によれば、「意味構造分野別語彙分析法」ということになるが、語彙研究の中で用いる場合に、一々「語彙」を付ける必要はないと思う。