

語彙

水谷静夫 著

朝倉書店

朝倉日本語新講座

2

語彙

水谷静夫 著

おぬ
て
み
く
じ
り
く

朝倉書店

著者略歴

1926年 東京に生まれる
1948年 東京大学文学部国文学科卒業
現在 東京女子大学教授

朝倉日本語新講座 2

語 彙

1983年4月5日 初版第1刷

著者 みず水 たに谷 しず静 お夫

発行者 朝 倉 邦 造

発行所 株式朝倉書店

東京都新宿区新小川町6-29

郵便番号 162

電話 03 (260) 0141

振替口座 東京 6-8673 番

<検印省略>

© 1983 <無断複写・転載を禁ず>

新日本印刷・渡辺製本

講座刊行の趣意

世に時々日本語ブームなるものが起る。その因となり果となるかのように日本語講座が既に幾つも現れている。自分たちの言葉への関心・反省が高まるのは喜ばしい。最近はまだ、電子計算機による日本語処理が盛んになった。単なる字面処理から進んで、文法・意味の情報を利用しつつ、言語表現に託せられた知識など内容面の操作にまで及ぶ勢いにある。

こうした時点で必要な事の一つは、日本語研究で得ている知見を処理技術とうまく結合することであろう。このための言語研究者側の努力は、理工系の人たちへの情報提供にとどまらず、自分らがして来た所を新たな角度から見直すことになり、どちらの側にも有益に違いない。そういう相互刺激の下に次代の日本語を考えるのも、時務として大切である。

この講座は、現代の書き言葉を中心に、問題点を整理しつつ日本語の姿を描き出そうとする。日本語も人類の言語の一つゆえ、日本語を論ずることで同時に言語一般に思いを致すという局面もあろう。時には言語哲学的な検討が必要になる。考察する表現の範囲は、技術文献・事務文書から文芸作品にわたるように努めたい。講座を次の通り組み立てる。

文字・表記と語構成	第一巻
語彙	第二巻
文法と意味	第三巻・第四巻
運用	第五巻・第六巻前半
人文系研究のための日本語処理入門	第六巻後半

要は、日本語の事實的・理論的な基礎知識の確認であり、従来の類書に比べては問題発見的・手続化志向の態度で臨むことを強調しておく。資料の部を所々(多くは巻末)に置くのも、この講座の特色となろう。

一つの講座である以上、全巻に用語・表記法まで統一の存するのが理想であろうが、そこまでには至らない。これも日本語の現実である。あるいはまた対立意見の並置を見るかも知れないが、それが研究の姿でもある。

序

この講座の刊行順で本書が露払いを勤めるが、他の巻と違って単独執筆である。これは次の考えによる。

語彙の論は、同時に念頭に置くべきものの数が格段に多い点で、音韻や文法の論より組織化がむずかしい。研究の歴史も浅い。泥臭い仕事を根気よく積み上げなければ成果は期し難く、たゞ根気にだけたよって方法的な見通しを欠けば收拾がつかなくなる。その上、現代日本語の語彙と言っても均質単一でなく^{ちよう}重疊し、部分的には流動している。その全貌を概観するには語彙論はまだ若過ぎる。にも拘らず日本語の様々な問題の解決に語彙論的考察も要することも、疑われない。そうした状況では、話題をとびとびに拾って述べるより、要求に応じた語彙論的情報をどんな仕方でも得るかに、かなり重きを置いて説くのが、かえって有用ではあるまいか。とすれば少なくとも現段階では、分担執筆にしない方が好都合かと考えられた。

さて普通の概論書なら、本書の第2章(資料作成法)や第3章(標本抽出法)は無くても済みますか、入れてもごく軽く触れるかする所である。そこに紙幅をかなり割いたのは、この種の配慮に無頓着な資料作りが跡を絶たないからである。電子計算機で大量の事例を調べたから安心だといったものではない。見通しを持して信頼度の高いデータを作り、一般性と特殊性との兼ね合いで適切に解析することが、大切である。日本語を説く講座の一冊でありながら本書に方法論的色彩が濃いのは、この理由による。話題の大半が計量語彙論の領域の事柄なのは、まさに語彙を対象とする語彙論(この事の意味は第1章に述べてある)の現情を反映すると言えよう。質的なものの数量化も始まっている。勿論、計量的でない語彙論が在ってよいばかりか、計量語彙論が語彙論のすべてではない。但し著者の見解では、言語の他の側面を扱う場合に比べ語彙的事象を扱う場合には、統計的な見方が必要でもあり有効でもある。

本書では解説例の多くを昭和前期の歌謡曲に採っているが、その趣旨を述べ

ておかなければ誤解を招くおそれがある。本書が歌謡曲用語の概観を目的としてはいないという事は、当然である。(もしそうなら、この程度のデータでは全く足りない。) 語彙論では、どんな分野の表現を取り上げるかの影響が現れやすい。使用率分布のように、表現内容がどうであれ(しかも日本語に限らず) 同じ型になると思われる事柄も在るが、多くは「新聞の場合こうだ」「学術誌の場合こうだ」という態の性急な一般化がしにくい。材料表現の世界に即した考察がしばしば要る。それだけに、材料表現の性格について予備知識を余り必要としないものを使ったかった。昭和前期歌謡曲に馴染みの薄い読者にも、歌謡曲というものの性格は容易に推察出来よう。割合短くて独立した表現でありながら極端には短くない点でも、語彙論の解説用に手頃である。

本書は、小さくても語彙論上の具体的意味を有つ事柄を解説例とし、解説例の積み重ねに日本語の語彙の姿を語らせるという、行き方を選んだ。例の間の脈絡にも相当注意したつもりである。抽象的に綺麗な話というのは、語彙論には無理である。また単に調査結果を並べ立てるだけでは、組織的な知見とはなり難い。結局の所こうした述べ方に落ち着いたが、諸例の背後に在るはずの何らかの一般性も汲み取って戴きたい。

1983 年春

水谷 静夫

歌謡曲歌詞の引用について、日本音楽著作権協会(出)許諾第 8213326 号。

目 次

1. 序 説	1
1.1 語彙論.....	1
1.2 計量語彙論.....	3
1.3 用語調査の理論的基礎.....	5
1.3.1 連 糸.....	6
1.3.2 語彙論の形式的前提.....	9
1.3.3 基本概念の定義.....	13
2. 用語データの作り方	17
2.1 用語調査.....	17
2.2 調査の企画.....	19
2.3 調査対象の定め方.....	22
2.4 本文のチェック.....	26
2.5 単位語・見出し語認定規則の作成.....	29
2.6 用語調査の全体的な流れ.....	34
2.7 作業方法の示唆.....	36
2.7.1 手作業に使うカード.....	36
2.7.2 電子計算機の利用.....	38
3. 標本抽出と使用率等の推定	43
3.1 平均・使用率および分散.....	43
3.1.1 平均と分散.....	43
3.1.2 度数分布表による平均・分散の計算.....	44
3.1.3 使用率も平均だ.....	47
3.2 抽出推定.....	47

3.2.1	標本抽出の考え	48
3.2.2	確率・期待値	49
3.2.3	区間推定法および誤差・偏り	53
3.3	基本抽出法による推定	55
3.3.1	有限母集団の場合の理論的考察	56
3.3.2	調査対象の平均値の推定例	59
3.3.3	無限母集団の場合	63
3.3.4	標本の大きさの決め方	64
3.4	層別抽出法による推定	65
3.4.1	層別という事	65
3.4.2	推定例と若干の注意	69
3.5	比推定	72
3.5.1	共分散・相関係数	72
3.5.2	比推定の考え方	76
3.5.3	延べ語数の比推定	79
3.5.4	使用率の比推定	81
4.	語彙に関する分布問題	86
4.1	どんな問題が在るか	86
4.2	語類構成比	90
4.2.1	大野の法則	90
4.2.2	樺島の法則, その他の問題	98
4.3	使用率の分布	102
4.3.1	使用率分布問題とは	102
4.3.2	種々の法則化	103
4.3.3	射影型分布函数の当てはめ	107
4.3.4	上位語による覆い率	113
4.4	語彙量推定問題	116

5. 共出現語に基づく解析	121
5.1 共出現語の定義.....	121
5.2 数量化Ⅲ類による用語・作品解析.....	123
5.2.1 Ⅲ類のキー項目の選び方.....	124
5.2.2 適用例.....	127
5.3 分割表数量化による仕訳.....	136
5.4 数量化Ⅳ類による作品解析.....	141
5.4.1 用語類似度の指標.....	141
5.4.2 適用例.....	143
5.4.3 双対的な問題.....	151
5.5 推移的閉包法など.....	152
6. 計量語彙論から見た文章展開	155
6.1 新出語率を追って.....	155
6.2 指標 D を使って.....	160
7. 基本語彙の問題	165
7.1 基本語彙とその類似概念.....	165
7.1.1 問題の在り処.....	165
7.1.2 概念規定の試み.....	168
7.2 基本度函数.....	172
7.3 用語網.....	178
附録 プログラム例	183
1. なまデータの整備.....	185
2. 基本抽出による平均値推定の教材.....	191
3. 語類構成比の線型回帰調整.....	193
4. 使用率分布函数の射影型近似.....	196
5. フィッシャの直接確率計算法.....	202

6. 萩野法の変種版	205
7. 指標 D の計算	210
参考文献	219
索引	223

図表目次

図		23-2	用語類似度 D 行列の色揚げ	149
1-1	調査の企画法	20	24	歌謡曲の行文を追った
1-2	企画図の一例	21		$n-k$ 関係
2	語彙調査の流れ	34	25	随筆の行文を追った
3	単位語カードの様式	37		新出語率
4	データファイル原稿	39	26	歌謡曲の行文を追った
5	整理作業の段取りの一例	41		指標 R
6	リフレーンと変異係数	46	表	
7	基本抽出法による推定例	59	1	昭和初期流行歌84篇
8	流行歌84篇の延べ語数分布	60	2	層別抽出法の推定結果
9	推定の試行列と信頼区間	62	3	比推定の推定結果
10	流行歌84篇の行数分布	69	4	古典の延べ語数と
11	層別抽出法の推定例	71		異なり語数
12	昭和初期流行歌の 行数と延べ語数	75	5-1	大野の法則(現代雑誌)
13	比推定(基本抽出)の一例	80	5-2	大野の法則(古典, その一)
14	古典の延べ語数と異なり語数	90	5-3	大野の法則(古典, その二)
15	大野による品詞別構成比	91	6-1	使用率分布函数の 射影型近似, その一
16	大野の法則(表 5-2 の場合)	99	6-2	使用率分布函数の 射影型近似, その二
17	樺島の法則	99	7	雑誌群の年間語彙量
18	使用率分布の射影型近似	107	8	梅・桜弁別キー語の値
19	古今集の梅・桜の歌の布置	128	9	歌謡曲とキー語との数量化
20	歌謡曲とキー語との数量化	134	10	分割表数量化による結果
21	キー語数量化の対照	140	11	歌謡曲の用語類似度 D
22	数量化 V 類による 歌謡曲仕訳	144	12	歌謡曲の数量化 V 類解析
23-1	V 類による歌謡曲仕訳, その二	148	13	随筆の新出語率
			14	基本度 f 上位七百語

1.1 語彙論

「語彙」という言葉は、案外、正確には使われない。ある言語に関し（その一定範囲の）あらゆる語を一まとめにして考えた総体が、語彙である。それはちょうど、個々人が人類でなく、個人を何万人か考えたとしてたゞそれだけのことなら「人類」とは呼べないようなものである。この言葉がかなりいい加減に使われる原因の一つは、「彙」が日常生活に馴染みの薄い文字で、従って意味がはっきりとはしないことであろう。

現存最古の辞書と言われる、中国の古辞書『爾雅』を調べると、この文字を次のように説いている。『爾雅』は分類体の辞書であるが、獣の名を集めたその積獣第十八に（十三経注疎本によれば）

彙毛刺 疎云、彙即蝟也、其毛如針

とあり、もともとハリネズミを指す字であった。ハリネズミの毛が沢山集まって生えているところから、アツメルという意味に転じ、アツメタモノ・アツマリの意味を生じた。

英語の“vocabulary”も、単に words のことではない。まずこの点に注意せられたい。一定の観点から捉えた《語の集まり》というところが大切である。

語彙の概念を正面からきちんと捉えて扱っては来なかったのは、言語学でも日常の用法と大差が無い。従来「語彙論」の名の下で研究していたのは、多くは、個々の語の形や意味、その変遷についての、記述であった。かような《語誌》を幾ら積み重ねても、それだけでは語彙論——語彙の論——とはならない。この点に関し『国語学大辞典』（1980、東京堂）の《語彙論》の項（執筆

者：樺島忠夫)は、こう書き起している：

音韻論・文法論と並べられる言語体系研究の一部門。言語の一つの単位である語を、ある規準または成立条件によって集合としてとらえたもの、すなわち語彙について、体系的に記述・説明する。ある時代・作品の中から語を任意に選びとって、その語源・用法・意味・価値を研究する論が語彙論を称することがあるが、これは語の論であって、語彙論ではない。〈中略〉語彙論は、語の集合を全体的に取り扱うものである。[354頁]

我が国だけでなく外国においても、この意味での語彙論は久しく現れるに至らなかった¹⁾。

では語誌的研究は語彙論には不要か。そんな事はない。個々の語の理解をゆるがせにするような態度で語彙の良い研究が出来るはずは無い。但し語誌研究と語彙論との縁の深さにもかかわらず、両者それぞれの目的は異なる。しかも語誌研究すら語彙の見地を離れては十分には進められない。例えば、近頃はその呼び名を余り使わなくなったが、「赤靴」という言葉がある。童謡の「赤い靴はいていた女の子」を持ち出すまでもなく、この語と「赤い靴」との意味するところは同じでない。一方「白靴」と「白い靴」、「黒靴」と「黒い靴」、「赤犬」と「赤い犬」等々、これらはそれぞれに、いわば同義である。なぜ靴の場合には異義なのか。そしてまた、やはり互に異義である「赤本」と「赤い本」との場合には、更に別の事情が存する。こうした事を明らかにする為には、たとい語誌研究であっても、この種の同類の形の語の対を集め、その集合を背景にして考究しなければならない。

今挙げた例のような事は、電子計算機による日本語情報処理の際の問題点の一つになる。類例は幾らも思いつく。同じく飲用の酒と湯とでも、「酒飲み」は人を、「湯飲み」は器物を指す(英語の接尾語 er/or にも似た事が起る)。この

¹⁾ 旧版の『国語学辞典』(1955)の同じ項目を読み比べると、そこには四半世紀の時の流れが感ぜられる。語彙の論であることを念頭に置いたものとしては、1950年に国立国語研究所で着手した用語調査が、世界的に言っても早い試みであった。語彙論のかような考え方を国語学界が受け容れ始めたのは、水谷静夫：語彙論の術語をめぐって、『国語学』[62] (1965) 71-84 あたりからだと言って過言ではあるまい。

種の事は機械だけでなく外国語習得の際にもしばしばつまづく所である。語彙についての広い見渡しを欠けば、その場当りの一時しのぎしか出来まい。先に説いた《語彙論》の必要な所以である。ところで語彙論は音韻論などとは違って、格段に多種の対象を扱わざるを得ない。適宜の語類（構文的カテゴリ）を利用して縮約を図る文法論に比べても、著しく不利である。方法論上のこの困難のゆえに、本来の語彙論はなかなか発達しなかった。この講座の一冊として方法論に重点の在る本書を著す理由は、まさにこゝにある。徒らな大量観察主義では、語彙の姿を明らかにし得ない。

1.2 計量語彙論

一つの言語に関する語の何らかのクラスという意味での語彙の構造的特色を計量的方法で追求することを主目的とするのが、計量語彙論である。その出現によって初めて語彙論に実質が与えられたとさえ言える。研究史上これは偶然とは思えない。

言語学の領域で一般に先進的分野と認められる音韻論に比べると、語彙論の発達は大変に遅れていた。一方、著者の見るところ、この両分野の理論的基礎は、次の節で述べる通り、形式的には大変よく似ている。にもかかわらず遅速の差が著しかったのは、それぞれが対象とするものの種類の多少に格段の開きがあるからだと思われる。音韻論が高々 10^2 のオーダーの対象を扱えばよいのに対し、例えば現代日本語の語彙の研究となると、少なくとも 10^4 、本格的にはそれより1桁か2桁か大きいオーダーの対象を考慮しなければならない。人間が一まとめに眺め得る範囲には限度があるから、この違いは重大である。語彙論の場合、そういう多種多様のものの全体を見渡す為に、集団的特性を簡潔に記述し、それに立って推測も行う手掛りとして、統計量を利用するに至った。語彙論が統計技法による計量語彙論を主流とするのは、だから自然の事である。

だが、そういう仕方で量的に扱うだけで質的な側面に着目しないでもいいのか——と、問い返されるかも知れない。実はこの間には、何を「量的」と言い何を「質的」と言うかを明確にした上でなければ、答えにくい。例えば今日我々

は速さを量的に見ることに慣れている。しかし事情は太古の人にとっても同じだったろうか。恐らく始めは、せわしなく脚を動かすアリの歩みと、ゆったり身を移すウシの歩みとを、直接的に思考の対象として比べることなど思いも及ばなかったかも知れない。単位時間当りの移動距離という考えの発明で、速さを、個々の身体的動作の形状から独立した量的特性として考え得るに至ったのではないか。量的特性は、それを認識する人間の側の発明に係る。徒らに手をこまねいて初めから対象の側に具わるものと考えてはならない。

本書で後に再び示す通り、普通には質的と見られる問題を計量的に解明した、次のような研究も在る。

古典文芸作品に現れた語の作品ごとの語類構成比に統計的な法則性の在ることを、大野晋が見出した。水谷静夫の追試によれば、これは現代の雑誌記事にも認められ、名詞類（体言）の構成比を説明変数として他の語類の構成比がかなり正確に推測出来る。こゝまではいわゆる質的な問題とは受け取られまい。ところでこの名詞類構成比について、その値の似たものの群れが、値の大きい方から順に、和歌・随筆、日記、物語というジャンル別に対応するという結果が見られた²。従来は専ら直観的に（そうでないにせよ言語表現以外の要因を考慮して）論ぜられたジャンル別に関し、言語表現のいわば内部から、しかも量的特性によって、これを論ずる手掛りが得られたわけである。

歌謡曲の世界では、ヒット曲が出るとその類似作品が輩出して《○○物》と一括される場合が多い。これら諸作の間に用語の共通度が高そうだとはい、誰しも思い着くところであろう。水谷が験した若干の事例では、この見込みが支持される。但し用語の共通性と言うだけでは甚だ漠としている。これを客観的に追求するにはそれなりの用意が要る。ここに活躍する方法として、林知巳夫の数量化理論（いわゆる第Ⅲ類、第Ⅳ類など）が在る。一つには適宜のキー語を選びそれらが使っているか否かのパターンを利用して、キー語ともども作品を数量化すること、一つには作品対の用語類似度を適宜に定義しそれに則って測った値により、作品間の似寄りを数量化すること；どちらにせよ、この数量化

² 大野の資料の作り方には問題が若干有るとの批判が出ている。それを修正してもここに記した事の大局は動かない。

の結果の布置が従前の《質的》判断を裏書きする成功例は多く、またそれらが喰い違った結果になった場合、却って数量化の結果の側から従来の見落としに気づくといった事も無いではない。水谷はこの方法を短歌の用語解析にも使って良好な結果を収めている。

歌謡曲の用語解析と言うと、多分に趣味的に響く。しかしそこで使った手法は、例えば科学技術文献の自動分類のようなものを考えても、原理的には同じはずである。本書では後で歌謡曲を例に取ることが多いが、それが割合に短くてしかもまとまった言語表現であるという点で、解説用には便利だからである。本書の目的が歌謡曲用語の研究に在るわけではない。

1.3 用語調査の理論的基礎

語彙論の知見は、その性格上、文法論などに比べると遙かに個別的である。よほど恵まれた場合でないと、他人の作ったデータがそのまま自分の関心に適合することにならない。そこで大概の場合には、データを得る為に用語調査に手を染めざるを得なくなる。しかも厄介な事に、同じ材料を対象範囲にしても、《語》の認め方如何でまるで違って見える結果となりやすい。大量に調べれば誤差は相殺するなど、気楽に考えては濟まない。その事情は第2章で次第に明らかになるが、語彙論の確度を高める心得の一つとして、用語調査に理論上どんな事が要請されるかを、まず見ておく必要が有ると思う。

かような考察は、語彙論自体と言うよりそのメタ理論に属する。今まで余り語彙論に馴染みの無かった方には、それほど分りよくあるまい。その場合にも一通りは目を通して戴きたい。幾つかの術語をこの部分で定義するし、何よりも、こうした方法的反省に無頓着に既成データを見るのは誤認に誘われ勝ちであり、また用語調査をしても徒勞になりかねないからである。

この節では、まず連糸 (string) に関する予備知識を述べる。それは、語彙論の直接材料が当の言語系 (language system) における連糸であって、そこから歩一步一步と操作を進め語彙の構造に至るはか無いからである。次に、通例期待される姿の語彙論を成り立たせるのに、少なくともどんな形式的前提が要るかを述べる。特に、ある程度大きい用語調査は標本抽出によるのが普通である

が、この点にも配慮して前提の要請をする。そしてこれに立って、語彙論の幾つかの基本概念を定義する。

1.3.1 連糸

人間は記号を使って生活している。言語も記号体系の一種である。と言うより《言語》を広く考えればどんな記号体系も言語だと見ることが出来よう。無論こゝではそれほどまで広くは考えずにおく。しかし我々は、言語記号について、経験事実として次の事を認めてよい：

現に使われた二つの記号の（任意の）組に関し、それらが同じものか否かが（形が単純な場合ならほとんど確実に）見分けられる。

例えば今述べた文で「現」の直後に書いたものと「組」の直後に書いたものとは《同じ平仮名》である。現れた場所が違うのだから、この一点をもってしても異なると言ってよいわけであるが、その事を認めた上で、この二つは（R. CARNAPの用語を借りれば）《意匠》を等しくすると受け取るのが普通である。そして見方のこの開きは token としての記号に着目するかその type に着目するかとの差である。

さて、一つの言語系に関しそれによる表現をある程度集めて眺めると、同じ意匠の反覆がある事が認められよう。しかも、集める表現の範囲を広くして行っても、意匠の種類（異なる意匠の数）には限度が有りそうである。こうした経験事実から我々は作業仮説として、

一つの言語には有限種の**基本記号**が有り、その言語表現はこの範囲でまかなわれる

と考えて進もう。「基本記号」と呼ぶからには、それが他の基本記号から合成されるものであってはならない³⁾とするのは自然であるが、こゝまでではまだ、どんなものを基本記号と見るかは定まらない。音素とか文字とか《単語》と

³⁾ 個々の漢字を基本記号と見るとしよう。「林」は「木」二つから成るから基本記号ではないのか。そうではない。「木」二つから成る——との表現が精密さを欠くのであって、もしこの論の通りなら「緑の林」と「緑の木木」との区別がつかない事になる。「林」は基本記号の「木」を横に並べたのでなく、それと形の似た構成要素（字原的に有縁だという事はこゝの論旨とは別の事）を横に並べたのである。基本記号が構成要素を有しても一向に差支えない。