

国立国語研究所学術情報リポジトリ

Collection of words from achral texts using the scout method

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 高梨, 信博, TAKANASHI, Nobuhiro メールアドレス: 所属:
URL	https://doi.org/10.15084/00001088

スカウト方式による用例採集法の 実験について

高 梨 信 博

1. はじめに
2. スカウト方式について
3. 実験の進めかた
 - 参加者
 - 資料
 - 採集の基準
4. 実験の結果
 1. 採集者によることなり
 - 採集量
 - 平均一致度
 - 2名の採集者による一致度
 - 低一致度語
 2. 資料によることなり
 3. 1回目と2回目の比較
 - 採集量
 - 同語採集の割合
 - 平均一致度
 - 語別的一致度
 4. 採集語について
5. おわりに

1. はじめに

昭和52年度にもうけられた国語辞典編集準備委員会をうけて、昭和54年度に創設されて以来、国語辞典編集準備室では、『日本大語誌』（仮称）の編集に関する諸問題を検討してきたが、その一環として、昭和57年12月および58年1月の2度にわたり、用例採集法に関する小実験をおこなった。

『日本大語誌』では、第1期の作業として、1901(明治34)年から1950(昭和25)年までの50年間を対象に、約3000万の用例を収集することが考えられているが、これを用例採集の方式からみると、ランダムサンプリングをくみあわせつつ、約1/3を総索引方式によって採集し、残る2/3をスカウト方式とよんでいる有意選択方式で採集することとなっている。これらの用例採集の方式のうち、総索引方式とランダムサンプリング方式については、方法論としての蓄積もみられるが、スカウト方式については、個々の採集者の判断という主観的な要因が含まれるため、方法論として言及するには困難な面がある。大規模かつ組織的な用例収集の実例の報告も、方法としての問題点にまで深くたちいったものはみることができない¹⁾。しかしながら、全体の2/3をスカウト方式で採集する『日本大語誌』の用例の質は、このスカウト方式によって左右されるところが大きいものと考えなくてはならない。スカウト方式による用例採集について、いちおうのめやすをえ、問題点をあきらかにするために小実験をこころみたゆえんである。

2. スカウト方式について

今回の実験についてのべるのにさきだち、スカウト方式という用例採集法について、略述しておきたい。

用例の収集にあたっては、条件がゆるせば、対象とする資料の総索引を作成することがのぞましいといえる。総索引方式の特徴は、資料に含まれるかぎりのすべての用語を網羅しうる点にある。それにともない、さまざまな視点に立った利用が可能になること、計量的な処理ができること、ある用語が存在しないことを確認しうることなどの利点がえられる。だが一方、総索引の作成に要するてまは膨大なものである。しかも、通常の言語資料の総索引によってえられる用語は、少数の高頻度語の割合が大きく、広い範囲にあらわれるいろいろな語をなるべく多く効果的に採集するためにはむだの多い方法といわなくてはならない。

この欠点をおぎなうため、できるだけ多くの資料から必要と判断した用語

のみを採集する方法が考えられる。これが本稿にいうスカウト方式である。スカウト方式においては、採否に関してある程度のめやすをもうけることはできるにしても、最終的には採集者の判断によって採否が決められることになる。スカウト方式の方法としての基本的な問題点は、原則として、この採集者の判断という要因によって生じるものということができよう。

総索引方式とスカウト方式の優劣は、比較の基準がかならずしも同じでない以上、単純には決しがたいが、要は総索引の作成に要するてまと、スカウト方式で結果的に生じざるをえない〈採集もれ〉とを、用例収集の全体計画のなかでどのように位置づけるかによることになるう。

ところで、資料を通読しながら必要と考える用例を採集してゆくというスカウト方式は、方法としてはけっして特別のものではない。特定の語や語群、または問題に応じて、資料から用例を採集するということは、多くの研究者がなんらかの程度でおこなっているはずである。だがそのような場合、多くは問題となる語に限られており、その限られた問題語を採集してゆくというやりかたでは、実は、採集者の判断は、あまりゆれを生じる余地がない。また、採集者も、ひとり、ないし十分に意思の統一をはかりうる程度の少数にとどまることが多い。

これに対して、『日本大語誌』の場合、一定期間内における日本語の全体像を語彙の面を中心に包括的にとらえようとするわけであり、採集にさきだつて、問題となる語をせまく限定するということとはできない。また採集者についても、かならずしも日本語研究の専門家とはいえない人々にも依頼せざるをえないであろう。そのような採集者に対して、事前に教育ないしは訓練といったことをおこなう必要があるとするならば、どのようなやりかたがもっとも有効であるかを明らかにすることが求められるわけである。

国語辞典編集準備室では、開設後まもない昭和54年度に、スカウト方式による用例採集法の実験をおこなっている。その実験は、個人間の採集のことなり解消するには、原資料について100語あたりどの程度の語数を採集すればよいかを知ることを目的としたもので、結局、100語あたり、半数の50

語以上を採集しなければ、個人間のことなりは消えないという結論がえられた²⁾。

しかし、総索引方式に対して、スカウト方式が用例採集方式としてもつ意味を考え、また、処理すべき資料の分量を検討してゆくと、のべ語数にして原資料の半数を採集するのでは、スカウト方式としての有効性にも問題を生じざるをえない。

こうした問題点をふまえて、国語辞典編集準備室では、見坊豪紀調査員が、140万枚におよぶカード採集の経験にもとづき、『スカウト式用例採集の手引き』(国語辞典編集準備資料8、昭和57・12)を執筆し、また夏目漱石の『坊つちやん』を資料に、さまざまなスカウト方式による用例採集をこころみ³⁾。今回の2度にわたる実験は、これらの成果にもとづいて実施されたものである。

3. 実験の進めかた

今回の実験は、昭和57年11月25日と昭和58年1月18日の2度にわたり、国立国語研究所でおこなった。

参加者 第1回目は13名、第2回目は15名であった。今回は、1・2回の共通参加者12名に限って報告する。12名のうちわけはつぎのとおりである。

記号	性別	年代	身分
A	男	30	所員
B	男	40	所員
C	男	60	準備室調査員(元所員)
D	男	60	準備室調査員(元所員)
E	女	20	準備室アルバイト
F	女	20	準備室調査員(大学院生)
G	女	20	準備室調査員(大学院生)
H	女	20	主婦
I	女	20	教員(国語担当)

J 女 30 銀行員
 K 女 50 主婦
 L 女 50 主婦

大きく二分すれば、A～Gの7名は国語辞典編集準備室に関係し、程度の違いはあれ、スカウト方式についての知識をもつものに対し、H～Lの5名は、まったくはじめてである。なお、学歴は、すべて大学または師範学校本科以上である。

資料 『用例採集のための主要雑誌目録』（国語辞典編集準備資料3）で選定された雑誌120誌のなかから、総合雑誌として「太陽」をとりあげた。今回予定されている用例採集の対象となる期間の第1年目にあたる明治34年の1月号から、分野・文体などの違いを考慮してページを選んだ（このページの選定は見坊調査員がおこなった）。実験においては、参加者によって採集の進度に遅速の差があったため、今回は、全員が採集を終えた部分にかぎって報告する。そのうちわけはつぎのとおりである。

<第1回>

- I 久保田譲：学政振張と財政（論説 文語） 6—7ページ 100行
- II 広津柳浪：櫛紅葉（小説雑俎 口語） 86—87ページ 88行
- III 芳川顕正：青年時代の苦学（上）（歴史地理 口語・談話筆記）
126—127ページ 100行
- IV 上野英三郎：煙草耕作地の制限・豊作と貯蓄・飼育界の警報（農業世界 文語） 166—167ページ 100行

<第2回>

- I 広告 乙の6—乙の7ページ
- II（第1回のIと同じ）
- III 坪谷善四郎：明治三十四年の経済界・昨年の経済問題（経済時評 文語） 62—63ページ 100行
- IV 麴町坊：印刷局女工
（無記名）：芸者の生存競争・東京語の勢力・都鄙の流行歌

(社会事情 文語) 204—206ページ 127行

V 鈴木東馬：欧米奇聞 (海外事情 文語) 206—207ページ 73行

ページと内容の切れめが一致しないために、ごく一部、上記以外の内容が含まれている場合がある。また、ここに示した行数は、採集数を比較するめやすとして、行取りにもとづいて算定したものであり、実際の行数とは若干ことなることがある。なお、実験では、各資料の内容の難易と採集のしやすさを考え、つぎのような順序で採集してもらった。

<第1回> II→III→I→IV

<第2回> III→IV→V→I→II

採集の基準 第1回目の際は、『日本大語誌』の計画の概要を説明したうえで、つぎのような基準で用例を採集してもらった。

- ① 自分自身にとって珍しいと思われるもの、現在とは違うと思われるものを採集する。
- ② 全数調査に対する補充なので、全数調査でも十分採集されそうなものはいらない。
- ③ 採集数は、1行あたり1個程度とする⁹⁾。
- ④ 採集すべき語の長さが決めにくときは長めにしるしをつけておく。

第1回目は、以上のようなごくおおまかな規定で採集してもらった。資料Ⅱの採集がおわった段階で、『スカウト式用例採集の手引き』(国語辞典編集準備資料8)の校正刷りのコピーを配布し、採集の着眼点について簡単な説明をおこなった。

第2回目の際は、実験の一週間ほど前に上記『手引き』をあらためて配付し、読んでおいてもらった。当日の指示は、原則として1回目と同じだが、1回目の反省にもとづき、つぎのような注意を加えた。

- ① 現代かなづかいとことなるかなづかい、および常用漢字表の字体とことなる字体については、それがいわゆる歴史的かなづかいや康熙字典体に一致するものであれば、現在とことなるものでも採集する必要は

ない。

- ② 用法上のちがいが無いと思えば、同じ語をくりかえして採集する必要はない。
- ③ 1回目は採集数のすくない人がいたので、1行あたり1個程度という採集量のめやすを守る。

なお、第2回目の実験では、Ⅱ（1・2回共通資料）について、

知らないことば

漢字の書きかた・用いかた

漢字の読みかた

意味・用法

などの区分によって採集理由を付記してもらったが、この点については、今回の報告では省略する。

なお、1・2回をとおして、採集に関する基準の説明等は、すべて見坊調査員がおこなった。

4. 実験の結果

1回目の資料Ⅰ～Ⅳ、2回目の資料Ⅰ～Ⅴのそれぞれについて、採集された語の五十音順語彙表を作成した。この語彙表では、それぞれの語を参加者のうちのだれが採集したかがわかるようにし、見出し語ごとに何人の参加者が採集したかを数え、その人数を一致度として付記した。

ところで今回の実験の結果をみると、われわれが前もって予想しなかったような問題がいくつかあらわれた。そのひとつは、第2回目の実験にさきだって採集者にあたえた指示にみられるように、いわゆる旧かなづかい・旧字体を採集した人がいたということである。これは、K・Lの2名がとくに多くおこなっており、旧かなづかい・旧字体で教育を受けたものにとっては、<今とは違うと思うもの>という基準で、こうしたものも想起されたのかもしれない。こうしたものまで語彙表に加えると、集計上、異質のもの加わっておそれがあるので、語彙表からは除いた。

もうひとつの問題点として、採集単位の長さがある。採集単位は長めにし
てよいという基準は、主として字音語などの長大な語形を予想したものであ
ったが、単語というよりは連語あるいは慣用句といったものの全体を採集す
るということがみられた。とくに文語文において、〈然らざることなからん
か・為す所を知らざるが如き・必要止むべからざる〉などをひとまとまりと
して採集したものがあつた。このような表現も、〈今とは違ふと思ふもの〉
に含まれるものと考えられたのかもしれないが、また一方では、こうした文
語の助詞・助動詞などがつらなつたものを単語にまで分割して採集するこ
とを求めるのはむずかしいということでもあろう。今回作成した語彙表では、
こうしたものは、他とのつりあいでも適宜見出しの長さを決めて調整した。

1. 採集者によることなり

はじめに、資料ごとの採集数と一致度の面から、採集者によることなりを
概観する。(表1～9)

表 1 採集者別の採集数と一致度 (第1回資料 I)

採集者	一致度												計	平均 一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	6	8	9	9	13	9	3	2	3	3	0	0	65	4.58
B	19	12	11	13	12	8	3	3	3	3	0	0	87	3.89
C	20	16	14	18	14	10	3	3	3	3	0	0	104	3.83
D	10	8	11	6	11	9	3	5	2	3	0	0	68	4.43
E	9	10	12	13	20	10	3	5	3	3	0	0	88	4.49
F	2	5	3	1	8	3	2	3	2	3	0	0	32	5.28
G	5	3	8	7	6	6	2	2	3	3	0	0	45	4.80
H	11	16	9	8	3	4	0	1	1	1	0	0	54	3.11
I	5	2	2	4	3	1	1	2	0	1	0	0	21	4.05
J	3	5	9	11	1	9	3	5	1	1	0	0	48	4.65
K	3	7	2	5	3	6	3	4	3	3	0	0	39	5.23
L	29	26	15	13	6	9	2	5	3	3	0	0	111	3.36
全 体	122	59	35	27	20	16	4	5	3	3	0	0	294	2.63
%	41.5	20.1	11.9	9.2	6.8	5.4	1.4	1.7	1.0	1.0	0	0	100.0	

表 2 採集者別の採集数と一致度（第 1 回資料Ⅱ）

採集者	一致度												計	平均一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	10	11	5	3	4	8	5	2	6	1	2	0	57	4.63
B	6	10	6	2	7	6	5	3	2	1	3	0	51	4.76
C	5	10	4	5	14	7	9	5	7	1	3	0	70	5.40
D	4	5	6	2	5	3	4	1	2	1	1	0	34	4.65
E	2	2	3	2	9	9	5	4	6	1	3	0	46	6.22
F	8	11	10	3	6	10	8	4	4	1	3	0	68	4.84
G	0	5	6	5	4	5	4	6	7	1	3	0	46	6.09
H	2	2	1	2	2	3	6	2	6	0	3	0	29	6.55
I	5	3	1	3	3	6	5	5	7	0	3	0	41	6.05
J	2	0	1	1	8	1	7	6	4	1	3	0	34	6.79
K	13	12	8	7	9	8	7	6	7	1	3	0	81	4.75
L	2	3	3	1	4	6	5	4	5	1	3	0	37	6.27
全 体	59	37	18	9	15	12	10	6	7	1	3	0	177	3.36
%	33.3	20.9	10.2	5.1	8.5	6.8	5.6	3.4	4.0	0.6	1.7	0	100.1	

表 3 採集者別の採集数と一致度（第 1 回資料Ⅲ）

採集者	一致度												計	平均一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	8	9	10	6	6	5	6	4	2	0	1	0	57	4.26
B	12	3	5	8	4	11	6	3	1	0	1	0	54	4.39
C	12	8	7	5	7	9	5	4	2	0	1	0	60	4.23
D	6	5	3	8	2	7	8	1	2	0	1	0	43	4.70
E	3	1	6	4	2	10	6	3	2	0	1	0	38	5.37
F	17	8	14	5	7	8	8	4	2	0	1	0	74	3.99
G	2	4	8	4	4	3	4	3	2	0	1	0	35	4.83
H	3	2	3	1	2	1	2	1	1	0	1	0	17	4.71
I	4	2	5	3	3	2	2	2	0	0	0	0	23	4.00
J	1	1	2	1	1	5	2	1	1	0	1	0	16	5.63
K	10	4	5	8	2	4	3	2	2	0	1	0	41	4.02
L	0	1	1	3	5	7	4	4	1	0	1	0	27	6.07
全 体	78	24	23	14	9	12	8	4	2	0	1	0	175	2.77
%	44.6	13.7	13.1	8.0	5.1	6.9	4.6	2.3	1.1	0	0.6	0	100.0	

表 4 採集者別の採集数と一致度 (第1回資料VI)

採集者	一致度												計	平均 一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	12	9	12	12	6	7	7	6	7	5	3	0	86	4.99
B	18	4	9	12	2	8	6	6	8	7	3	0	83	5.11
C	14	18	19	17	11	6	10	8	9	7	3	0	122	4.79
D	7	8	5	6	6	5	8	5	8	6	3	0	67	5.72
E	3	11	11	8	6	9	9	7	9	7	3	0	83	5.75
F	10	10	16	15	9	9	10	7	9	7	3	0	105	5.21
G	2	1	8	7	3	5	7	5	5	5	3	0	51	6.24
H	4	4	6	3	4	3	1	4	3	4	3	0	39	5.62
I	2	5	7	3	1	1	0	4	4	6	2	0	35	5.89
J	3	2	1	3	0	2	2	0	3	4	2	0	22	6.23
K	2	7	5	7	7	5	9	5	9	5	2	0	63	6.03
L	3	11	12	3	5	6	8	7	7	7	3	0	72	5.72
全 体	80	45	37	24	12	11	11	8	9	7	3	0	247	3.35
%	32.4	18.2	15.0	9.7	4.9	4.5	4.5	3.2	3.6	2.8	1.2	0	100.0	

表 5 採集者別の採集数と一致度 (第2回資料I)

採集者	一致度												計	平均 一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	8	8	12	19	7	13	9	12	10	6	3	4	111	5.76
B	13	16	21	17	8	16	14	14	9	5	3	4	140	5.24
C	4	8	7	12	8	11	11	12	8	5	3	4	93	6.14
D	4	3	4	3	6	6	8	11	7	5	2	4	63	6.78
E	5	9	9	20	9	11	11	12	8	6	3	4	107	5.86
F	0	2	5	11	7	13	7	4	5	5	3	4	66	6.56
G	3	4	11	10	9	16	12	15	8	6	3	4	101	6.36
H	4	13	6	11	5	12	10	11	4	2	3	4	85	5.69
I	24	11	12	12	7	8	8	9	5	4	2	4	106	4.68
J	11	7	6	5	6	10	9	8	8	5	3	4	82	5.89
K	10	20	17	16	17	15	12	15	8	5	2	4	141	5.22
L	17	9	19	16	16	19	15	13	10	6	3	4	147	5.37
全 体	103	55	43	38	21	25	18	17	10	6	3	4	343	3.62
%	30.0	16.0	12.5	11.1	6.1	7.3	5.2	5.0	2.9	1.7	0.9	1.2	99.9	

表 6 採集者別の採集数と一致度 (第 2 回資料Ⅱ)

採集者	一致度												計	平均一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	2	3	5	2	5	11	9	1	11	10	5	3	67	7.21
B	8	5	5	7	8	15	14	2	13	10	5	3	95	6.45
C	14	17	12	9	12	15	10	3	11	10	5	3	121	5.40
D	10	1	3	5	5	8	7	1	11	7	5	3	66	6.53
E	4	6	7	7	9	13	13	2	14	10	5	3	93	6.58
F	0	3	1	2	3	3	6	2	8	9	5	3	45	8.02
G	9	14	9	7	8	10	8	3	10	9	5	3	95	5.72
H	1	6	4	7	5	7	7	3	11	11	3	3	68	7.07
I	3	2	4	3	7	8	12	3	10	9	5	3	69	7.14
J	2	12	3	5	4	10	11	0	10	5	3	3	68	6.19
K	7	2	6	4	9	10	9	3	15	11	4	3	83	6.77
L	22	13	13	10	5	10	13	1	11	9	5	3	115	5.14
全 体	82	42	24	17	16	20	17	3	15	11	5	3	255	3.86
%	32.2	16.5	9.4	6.7	6.3	7.8	6.7	1.2	5.9	4.3	2.0	1.2	100.2	

表 7 採集者別の採集数と一致度 (第 2 回資料Ⅲ)

採集者	一致度												計	平均一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	7	9	7	9	12	9	11	8	13	7	12	3	107	6.47
B	10	10	7	12	12	13	8	12	14	5	13	3	119	6.25
C	5	12	5	12	11	15	10	9	13	7	13	3	115	6.48
D	5	8	8	8	9	6	6	3	7	4	9	3	76	6.08
E	0	1	6	11	11	9	9	14	12	7	13	3	96	7.24
F	0	1	0	6	9	8	7	10	10	3	12	3	69	7.77
G	4	9	8	4	9	13	9	9	7	7	12	3	94	6.61
H	8	4	8	13	12	12	9	9	9	6	12	3	105	6.36
I	8	4	11	6	13	9	8	9	10	5	12	3	98	6.41
J	8	16	11	17	11	8	3	7	8	6	10	3	108	5.55
K	11	18	16	19	18	11	11	10	11	6	13	3	147	5.59
L	10	18	15	15	18	13	7	12	12	7	12	3	142	5.70
全 体	76	55	34	33	29	21	14	14	14	7	13	3	313	4.08
%	24.3	17.6	10.9	10.5	9.3	6.7	4.5	4.5	4.5	2.2	4.2	1.0	100.2	

表 8 採集者別の採集数と一致度 (第2回資料IV)

採集者	一致度												計	平均一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	9	12	6	8	6	9	13	11	7	17	7	5	110	6.51
B	13	15	16	13	7	4	13	13	8	16	6	5	129	5.85
C	9	13	12	6	14	9	15	14	8	17	7	5	129	6.32
D	10	8	11	7	10	4	11	10	5	14	7	5	102	6.27
E	4	11	12	3	12	7	13	15	8	17	7	5	114	6.72
F	0	0	1	0	7	1	5	7	7	11	4	5	48	8.58
G	6	14	10	4	5	7	11	11	7	12	7	5	99	6.38
H	2	4	5	4	6	4	10	14	3	11	6	5	74	7.30
I	8	8	6	2	9	1	9	9	3	15	7	5	82	6.71
J	7	12	13	7	13	3	10	11	7	15	6	5	109	6.22
K	8	6	6	8	9	4	11	12	5	17	6	5	97	6.73
L	8	11	13	6	7	7	12	9	4	18	7	5	107	6.35
全 体	84	57	37	17	21	10	19	17	8	18	7	5	300	4.00
%	28.0	19.0	12.3	5.7	7.0	3.3	6.3	5.7	2.7	6.0	2.3	1.7	100.0	

表 9 採集者別の採集数と一致度 (第2回資料V)

採集者	一致度												計	平均一致度
	1	2	3	4	5	6	7	8	9	10	11	12		
A	4	8	4	5	7	3	2	4	5	3	0	1	46	5.17
B	12	8	8	9	3	2	3	4	5	3	0	1	58	4.40
C	11	5	10	7	6	2	3	4	4	3	0	1	56	4.50
D	4	3	4	4	4	1	2	1	2	2	0	1	28	4.86
E	5	6	8	8	4	1	3	3	5	3	0	1	47	4.91
F	0	0	0	1	0	0	0	1	4	1	0	1	8	8.75
G	2	5	2	7	6	2	1	3	5	3	0	1	37	5.57
H	3	4	5	3	1	1	0	3	3	3	0	1	27	5.26
I	3	2	1	5	3	1	0	2	2	1	0	1	21	5.10
J	7	5	6	8	2	3	2	2	4	3	0	1	43	4.72
K	3	0	2	1	1	0	2	3	2	2	0	1	17	6.24
L	10	8	10	6	3	2	3	2	4	3	0	1	52	4.27
全 体	64	27	20	16	8	3	3	4	5	3	0	1	154	2.86
%	41.6	17.5	13.0	10.4	5.2	1.9	1.9	2.6	3.2	1.9	0	0.6	99.8	

採集量 1行あたり1個程度というめやすを守ることは、やさしくはない。とくに、第1回目では、ほとんどの採集者の採集が1行あたり1個より、かなりすくなかった⁴⁾。2回目は、このめやすを強調したためもあって、1行あたり1個程度にちかづいているが、逆にかなり大きく超過したものもある。採集数は、どのような用例を採集するかという内容面からの基準にもとづいて採集をおこなった結果であって、一応のめやすにはできて、あまり強い拘束力はもちえないようである⁵⁾。

平均一致度 ひとりの採集者が採集した語例の一致度(12名中、何名が採集したか)の平均である。一般的には、この平均一致度が低ければ、他の採集者とはことなる独自の判断にもとづく採集がなされ、平均一致度が高ければ、他の多くの採集者と共通の判断による採集がなされているものと考えられる。もちろん、平均一致度の高低は、そのみでは、個々の採集者の評価といったものと結びつくものではない。今回の実験の範囲では、あくまで、どのような分布をなすかを知るためのものである。

表1～9の平均一致度をみると、平均一致度は、各採集者の採集数と関連するように思われる。たとえば、表1から平均一致度をとりだして大きい順にならべ、その下に実際の採集数を付記すれば、表10ようになる。

表 10 平均一致度と採集数

順位	1	2	3	4	5	6	7	8	9	10	11	12
平均一致度	5.28	5.23	4.80	4.65	4.58	4.49	4.43	4.05	3.89	3.83	3.36	3.11
採集数	32	39	45	48	65	88	68	21	87	104	111	54
採集者	F	K	G	J	A	E	D	I	B	C	L	H

採集数との比較でいうと、IとHの平均一致度がかかなり低く、逆にEの平均一致度はやや高いが、全体としては、採集数と平均一致度のあいだに関連が認められる。この傾向は、表2～9についても同様であり、それぞれ、若干の例外を含みながら、採集数のすくないものほど平均一致度が高くなっている。

このことは、今回の実験程度の採集数の範囲内では、採集数がすくないほ

ど、全体に共通する用例を採集しやすいということを意味する。いいかえれば、この段階では、個人的なカタよりは、あまり強くは出ず、その資料のなかの、特定少数のめだつた語形に採集が集中するということであろう。

2名の採集者による一致度 さきの表に示した一致度は、12名の参加者全員によるものであったが、もし標準的な採集というものが設定できるとすれば、その標準的な採集と比較することによって、各人の採集の状況を評価することができよう。実際には、採集には、客観的にみて絶対的な基準があるわけではなく、標準的な採集といったものを設定することは困難であるが、今回の実験で、採集の方針を定め、『スカウト式用例採集の手引き』を執筆した見坊調査員の採集を比較の基準とし、その他の採集者がどれだけ一致しているかをみておくこととしたい。

この一致度の比較としては、まず、見坊調査員の採集した語のうちのどれだけを各採集者が採集したかをみるべきことはいうまでもないが、今回の実験のように、参加者による採集数のばらつきが大きい状態では、上記の比率のみでは十分とはいえない。そこで、あわせて、各参加者の採集総数のうちにしめる見坊調査員の採集と一致した語の割合も示すこととする。たとえば、前者は命中の実数であり、後者は命中率といえるだろう。(表11・12)

はじめに、各人の採集語のうちの何%が見坊調査員の採集と一致しているか(表11・12の④)をみると、資料によって若干のことなりがあるが、全体の平均では、1回目が52%、2回目が55%で、半数強が一致することが知られる。この割合は、各人の採集数と相関するのではないとも考えられるが、今回の実験では、そうした傾向は、とくに明確にはあらわれていない。

つぎに、各人の採集語が見坊調査員の採集した語の何%をおおっているか(表11・12の③)をみると、さきにも述べたように、各人の採集数にかなりばらつきがあるため、単純には比較できないが、全体では、1回目が31%、2回目が44%となっている。これは、とくに1回目の実験で、見坊調査員以外の参加者の採集数が見坊調査員の採集数をかなり下まわっているため

表 11 各採集者と見坊調査員との一致度 (第 1 回実験について)

資料 採集者	I				II				III				IV				全 体			
	①	②	③	④	①	②	③	④	①	②	③	④	①	②	③	④	①	②	③	④
A	65	36	35	55	57	26	37	46	57	30	50	53	86	50	41	58	265	142	40	54
B	87	39	38	45	51	28	40	55	54	23	38	43	83	46	38	55	275	136	38	49
(C)	104	104	100	100	70	70	100	100	60	60	100	100	122	122	100	100	356	356	100	100
D	68	24	23	35	34	14	20	41	43	18	30	42	67	48	39	72	212	104	29	49
E	88	46	44	52	46	32	46	70	38	23	38	61	83	54	44	65	255	155	44	61
F	32	15	14	47	68	32	46	47	74	29	48	39	105	65	53	62	279	141	40	51
G	45	23	22	51	46	31	44	67	35	14	23	40	51	36	30	71	177	104	29	59
H	54	12	12	22	29	18	26	62	17	5	8	29	39	18	15	46	139	53	15	38
I	21	7	7	33	41	27	39	66	23	8	13	35	35	26	21	74	120	68	19	57
J	48	20	19	42	34	27	39	79	16	8	13	50	22	10	8	45	120	65	18	54
K	39	20	19	51	81	43	61	53	41	15	25	37	63	43	35	68	224	121	34	54
L	111	32	31	29	37	21	30	57	27	18	30	67	72	46	38	64	247	117	33	47
平均			24	42			39	58		29	45			33	62				31	52

表 12 各採集者と見坊調査員との一致度（第 2 回実験について）

資料 採集者	I			II			III			IV			V			全 体		
	①	②	③ ④	①	②	③ ④	①	②	③ ④	①	②	③ ④	①	②	③ ④	①	②	③ ④
A	111	50	54 45	67	47	39 70	107	61	53 57	110	73	57 66	46	27	48 59	441	258	50 59
B	140	59	63 42	95	57	47 60	119	65	57 55	129	78	60 60	58	31	55 53	541	290	56 54
(C)	93	93	100 100	121	121	100 100	115	115	100 100	129	129	100 100	56	56	100 100	514	514	100 100
D	63	34	37 54	66	43	36 65	76	38	33 50	102	64	50 63	28	17	30 61	335	196	38 59
E	107	47	51 44	93	58	48 62	96	62	54 65	114	72	56 63	47	22	39 47	457	261	51 57
F	66	31	33 47	45	31	26 69	69	44	38 64	48	39	30 81	8	6	11 75	236	151	29 64
G	101	52	56 51	95	54	45 57	94	55	48 59	99	58	45 59	37	24	43 65	426	243	47 57
H	85	28	30 33	68	45	37 66	105	61	53 58	74	53	41 72	27	14	25 52	359	201	39 56
I	106	32	34 30	69	45	37 65	98	49	43 50	82	52	40 63	21	11	20 52	376	189	37 50
J	82	32	34 39	68	31	26 46	108	48	42 44	109	66	51 61	43	17	30 40	410	194	38 47
K	141	53	57 38	83	55	45 66	147	70	61 48	97	62	48 64	17	8	14 47	485	248	48 51
L	147	57	61 39	115	57	47 50	142	68	59 48	107	62	48 58	52	19	34 37	563	263	51 47
平均	46	42		48	61		49	54		48	65		32	53		44	55	

- * ① 各採集者の採集数
 ② ①のうち、見坊調査員の採集と一致するもの数
 ③ ②÷見坊調査員の採集数(×100)
 ④ ②÷①(×100)

であり、もし見坊調査員と同数を採集すれば、上述のとおり、全体では各人の採集語の過半数、すなわち見坊調査員の採集した語の過半数ちかくは採集しうるものと推定される。

低一致度語 一致度の高い用語は、多くの採集者が共通して着目したものである。これに対し、低一致度語とくに一致度1の語は、少数の採集者のみが着目したものであり、個々の採集者の特徴を示す部分のひとつである。スカウトの効率という点からみても、低一致度語にあたる部分で、どれだけ有効性の高い用例を採集できるかが、採集者のよしあしを決定する要因のひとつとなる。

第1回目の実験から、資料Iについて、一致度1となったものを採集者ごとに例示すれば、つぎのとおりである。

- A (6項) 研究 公債 国務 時機 非常 茫然
 B (19項) 教育 教育費 教員 経済 高等学校 国政 財政 財力
 仔細 師範学校 司法 資力 精密 政務 大学 躊躇
 到底 暴論 陸軍
 C (20項) あらざるか 一策 学政振張問題 学政全般 如之 関係
 区域 教育事業 行政 減ず 事実 奨励補助 数百千万
 円 前代 第一方策 大成 第二策 帝国 二大方策 百
 分の五 やいな
 D (10項) 或は 異議 一旦 嫌 国勢 加之 実状 何程 果して
 不当
 E (9項) 充つ 及 各般 格別 己に と 兵備 寧ろ をして
 F (2項) 競ふ 来る
 G (5項) 拡張 議会 断行 単に 立論

- H (11項) あらざるべし 確立 過大 計画 期す 答ふ 財政当局
者 せざるべし 対す 若くは もの
- I (5項) 一事 先進諸国 何ぞや 下 故を以て
- J (3項) 予め 否 放擲
- K (3項) 境遇 図る 依る
- L (29項) 明か 非ず あらん あり 未だ 得ず 顧みる 感ず
ざるべからず 而 して 状況 すべし 為め 所 なか
るべし 何人 二箇 のみ べし 外 増す 免れる 認
む 見る 以て 基く 余裕 >

たとえば、Bではすべて漢語であるのに対し、Lではほとんどが和語である。こうした対立は、一致度1の語の個人差をなす要因のひとつである。また、この資料では、H・Lを中心に、文語的表現といったものが一致度1となっていることが知られる。

これ以外の資料では、人名・地名・書名などの固有名詞を採集するかいなか、助詞・助動詞などを採集するかいなかなどで、一致度1という個人的特徴のうかがわれるものがある。採集者個個人の特徴は、一致度1の語においてのみあらわれるものでないことはいうまでもないが、各採集者の採集の傾向をつかみ、適切な指示を与えるには、こうした視点からの観察も必要であると思う。

2. 資料によることなり

スカウト方式による用例採集は資料の通読を前提とするものであり、したがって、採集者の判断と同時に、資料のがわの性格にも影響されるところがあるものと考えられる。スカウト方式の場合、どのような資料が用例採集をおこないやすいものであるかということは、個々の採集者によってことなるところもあるが、一般的には、

- ① 目標とする採集数に容易に達しうること。
- ② 採集すべき用例の選択が容易であること。

の2点があげられる。今回の実験でいえば、①は採集数であり、②は参加者

全体の問題としてみれば、全体としての平均一致度におきかえることができよう。今回の実験結果にもとづき、資料ごとに、参加者1名あたりの平均採集数と、全体の平均一致度を示せば、表13のようになる（平均一致度の高い順に示す）。

表 13 資料別の平均一致度と平均採集数

平均一致度	4.08	4.00	3.86	3.62	3.36	3.35	2.86	2.77	2.63
平均採集数	106	100	82	104	50	69	37	40	64
資 料	2-Ⅲ	2-Ⅳ	2-Ⅱ	2-Ⅰ	1-Ⅱ	1-Ⅳ	2-Ⅴ	1-Ⅲ	1-Ⅰ

今回の実験では、採集数を1行あたり1個程度と限定しており、採集数は、資料による採集のしやすさを判断する材料にはなりえない。1行あたり1個程度という指示にもかかわらず、とくに1回目では、採集数がすくないが、これは資料の性格よりも、スカウトに対するふなれの方が大きく作用しているものと考えらるべきであろう。

採集すべき用例数が限定されてしまえば、資料によるスカウトのしやすさは、上記の二つの要因のうち、②に集中することになる。だが、表から知られるとおり、参加者全員による各資料ごとの平均一致度は、平均採集数と相関しており、平均採集数の均一でない今回の実験では、この平均一致度をもって、ただちに資料ごとのスカウト採集のしやすさを示すものとはできない。

結局、今回の実験では、条件として一定にすべき要因が変動してしまったために、資料の性格によるスカウトのしやすさということは、客観的に測定することはできない。しかし、実験後の参加者の感想でも、資料による採集のしやすさには違いがあるものと考えられる。内容の難易、文語文と口語文といった文体のちがひ、漢語含有率の大小、専門的用語の有無、ふりがなの有無などの諸条件によって、どのようにスカウトのしやすさが変わるものか、今回の実験からは、すべて今後の課題として指摘するにとどめざるをえない。

3. 1回目と2回目の比較

スカウト方式による用例採集は、採集者の判断という主観性の強い基準に

もとづいてなされるが、おのおのの採集者にとっても、自分の判断基準がつねに一定しているわけではあるまい。同一の資料であっても、時その他の条件をことにすれば、採集は変わってくるものと思われるが、それはどの程度のものであろうか。このような問題を考えるために、1回目と2回目の実験に、まったく同一の資料を加えてみた⁹⁾。

採集量 採集量は指定しているので、本来ならば違いはないはずだが、G・I・Kの3名で2回目の採集量がとくにいちじるしくふえている。このほかのものも、Dが2語すくなくなったほかは全員多少ずつふえている。

同語採集の割合 採集数にはさほどの違いがなくても、採集の内容まで同じとは限らない。1回目と2回目の採集を比較し、同じ語をどのくらいの割合で採集しているかを表にして示せば表14のようになる。

表 14 同語採集の割合

採集者	1回のみ 採集	%	1・2回と も採集	%	2回のみ 採集	%
A	26	28	39	42	28	30
B	27	29	41	44	25	27
C	29	19	75	50	46	31
D	24	20	63	53	32	27
E	39	30	49	37	44	33
F	21	32	11	17	34	52
G	20	17	25	22	70	61
H	44	39	10	9	58	52
I	16	19	5	6	64	75
J	19	22	29	33	39	45
K	18	18	21	21	62	61
L	77	40	44	23	71	37
平均	30	27	34	30	48	43

2回目の採集数の方が多い採集者にとっては、2回のみ採集した語の割合が大きくなるのはやむをえないが、1回目と2回目の採集数がほぼ同じものでも、1・2回とも共通して採集した語は多くはなく、2割から5割程度である。この数値は、2回の実験においてえられたものにすぎず、スカウトになれることによって、このようなゆれも小さくなるのではあろうが、やは

りそれは限界があり、個々の採集者の内部でのゆれは、かなりの程度で残らざるをえないのではないかとと思われる。

平均一致度 12名の参加者とも、2回目の方が平均一致度はかなり高くなっている。これは、各人の採集数との相関をこえるものである。とくに、12名の延べ採集数は2回目の方が200あまり多いにもかかわらず、ことなりでは逆に2回目の方が40ほどすくない。1回目のふなれによるばらつきが減じ、2回目には、特定の語への集中をましたものと考えられる。

語別の一一致度 上述のように、全体的な平均一致度では1回目よりも2回目の方が高くなっているが、採集された個々の語についてみれば、一致度の高くなった語もある一方で、逆に低くなったものもみられる。これらの一致度の変化は、最初に目につきやすいが、のちには不要と考えられたもの、はじめは気がつきにくい、のちには一致して重要と考えられたものなどといった性格を反映するものと思われる。

1・2回目のうち、いずれか一方での一致度が0のもの、つまりだれも採集しなかったものも含めて、一致度の変化ごとに項目数をまとめた表を作れば、表15ようになる。

表 15 一致度の変化

一致度 の変化	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	(計)
項目数	1	0	1	6	17	31	79	50	58	33	25	22	9	9	7	1	1	350

* 一致度の変化とは、各語ごとに、2回目の実験での一致度から、1回目の実験での一致度を引いた値である。したがって、たとえば一致度の変化が0ということは、その語の一致度が1・2回目とも同じであったことを意味する。

一致度の変化は、-12から+12までありうるはずであるが、実際には、-7から+9の範囲である。このうち一致度の変化の大きいもの、-3以下と+3以上のものについて、具体例をあげれば、つぎのごとくである。

- 7 (1項) 痛く
- 5 (1項) 左れば
- 4 (6項) 暇あらず が 証明 須らく 大 端緒

- 3 (17項) 挙ぐ 至る 況や 概ね 如し 施設 総括 当然 な
くんばあらず なさしむ なし の 判断し来る 百般
許す 余地 論ず
- + 9 (1項) 師範学校
- + 8 (1項) 放擲
- + 7 (7項) 官立学校 軽重緩急 裁判所構成法 実業学校 重複無
用 適當公平 文部
- + 6 (9項) 学制調査会 教育事業拡張 国語調査会 国政百般 歳
計 振興拡張 先進諸国 妄進 幼年学校
- + 5 (9項) 区々 公債 奨励補助 政務 戦後経営 大体上 秩禄
処分 地方公共団体 通信
- + 4 (22項) 学政振張 過大無益 緩急軽重 教育事業 行政諸般
緊急必須 決す 建議 高等学校 加之 私人 推移変
転 政費 大学 長距離電話 帝国教育 伝染病研究所
内外 二大方策 兵備 方今 漫然
- + 3 (25項) 一大計画 永楽病院 解釈 各局 各省 学政 学政振
張問題 洪大 国家歳入 財政 財政当局者 司法 首
尾一貫 商量精査 資力 前代 増設 多事 提議 内
務 八年計画 目論見 立論 隆替 吾輩

一見してあきらかなように、一致度の減じているものには和語が多く、一致度の高まっているものには漢語、とくに変化の大きいものでは3字以上の漢語が多い。このような傾向は、例を引くことは省略したが、一致度の変化が-2~+2のあいだの語でも、ほぼ同様である。

4. 採集語について

採集された用例の価値は、ある程度の用例数が得られて、問題の所在があきらかになった段階でなければ、簡単には云々できない。しかし、スカウト方式による用例採集の有効性を考えるには、やはり採集された一つ一つの用例に対する評価をぬきにするわけにはいかない。そこで、本稿では一つのこ

ころみとして、『日本国語大辞典』（以下『日国大』と略称する）を比較の対象とし、つぎのような点から、採集された用例の評価をおこなってみたい。

- ① 『日国大』に見出しのない語。
- ② 『日国大』に見出しはあるが、用例のない語。
- ③ 『日国大』に用例があるが、さらに用例を追加すべきであると考えられる語。初出例となりうるもの、その表記の例があげられていないもの、辞書などからの引用のみで、具体的な文脈内における使用例が示されていないものなど、比較的、少数に限る。

なお、多義語の場合は、採集された用例が該当する『日国大』の意味区分ごとに判断する。

以上の区分に従い、各資料ごとに該当する項目数を求め、一致度によって分類し、表にまとめれば、表16・17のようになる。

表 16 『日国大』との対照結果（第1回実験について）

資料	一致度 区分	1	2	3	4	5	6	7	8	9	10	11	12	計
		I	①	13	9	3	1	6	2	2	0	1	2	
②	1		0	1	0	0	0	0	0	0	0	0	0	2
③	3		1	4	2	2	0	0	1	0	1	0	0	14
II	①	3	0	3	0	1	0	1	1	1	0	0	0	10
	②	0	0	0	0	0	0	0	0	0	0	0	0	0
	③	5	0	0	1	3	3	4	3	2	0	1	0	22
III	①	5	2	2	2	3	5	1	1	1	0	0	0	22
	②	0	0	2	1	1	0	0	0	1	0	0	0	5
	③	4	1	2	0	1	0	1	0	0	0	0	0	9
IV	①	5	10	7	8	3	6	3	5	6	5	2	0	60
	②	0	2	2	1	2	2	4	1	1	2	0	0	17
	③	5	3	2	5	2	0	0	0	1	0	1	0	19
全体	①	26	21	15	11	13	13	7	7	9	7	2	0	131
	②	1	2	5	2	3	2	4	1	2	2	0	0	24
	③	17	5	8	8	8	3	5	4	3	1	2	0	64
計		45	28	28	21	24	18	16	12	14	10	3	0	219

表 17 『日国大』との対照結果（第2回実験について）

資料	一致 区分	1	2	3	4	5	6	7	8	9	10	11	12	計
		I	①	14	12	10	9	7	10	6	12	5	3	
	②	6	1	3	3	1	2	1	0	1	1	0	0	19
	③	3	1	2	0	0	1	0	2	1	0	1	1	12
II	①	3	5	5	3	3	6	5	2	3	2	3	4	44
	②	0	1	0	0	0	0	0	0	0	1	0	0	2
	③	1	1	1	0	1	2	1	1	2	2	0	0	12
III	①	2	2	4	6	5	6	7	6	3	2	6	3	52
	②	2	0	2	0	1	1	0	0	2	0	1	0	9
	③	0	2	0	1	0	2	1	0	0	0	0	0	6
IV	①	5	2	4	2	2	3	5	7	4	11	3	3	51
	②	1	0	0	0	1	0	0	0	0	2	0	1	5
	③	1	1	1	2	1	0	2	2	3	1	2	0	16
V	①	1	0	4	2	1	1	2	4	1	2	0	0	18
	②	0	1	0	0	0	0	0	0	0	1	0	0	2
	③	5	0	1	1	2	0	0	0	1	1	0	1	12
全体	①	25	21	27	22	18	26	25	31	16	20	13	12	256
	②	9	3	5	3	3	3	1	0	3	5	1	1	37
	③	10	5	5	4	4	5	4	5	7	4	3	2	58
計		44	29	37	29	25	34	30	36	26	29	17	15	351

1 回目の実験の資料 I から採集された語で、『日国大』に見出しのないものを例示すれば、つぎのようなものである。

一小局部 一大計画 一大決心 一も 永楽病院 学税 学政振張 学政振張問題 学政全般 学制調査会 過大無益 緩急軽重 関係区域 教育事業 緊急必須 軽重緩急 国語調査会 財政当局者 歳入総額 商量精査 奨励補助 振興拡張 推移変転 数百千万円 噬臍の悔 戦後経営 先進諸国 第一方策 大体上 第二策 秩禄処分 重複無用 直轄事業 帝国教育 適當公平 二大方策 判断し来る 百分の五 不可能的

一見してあきらかなように、3字以上、とくに4字の漢語が多い。そのほかのものも、『日国大』の立項の方針からみて、当然、見出しにはなりえないものが多い。したがって、これらがすべて辞書の見出しに立てるべきものであるというわけではないが、すくなくとも用例その他のかたちで記録に残しておく必要はある⁹⁾。

このほかの資料から採集された語のうち、歴史的な辞書の見出しとして立ててもよいのではないかと思われるものには、つぎのようなものがある。

含液 牛畜 固性 成畜 双蹄獣 致死症 倒用 毒力 肉乳 農業者
肥糞 病機 幼畜 (以上、第1回資料Ⅳより)

一試 会費金 奇剤 顧問医 心育 送荷 大盛 呈送 薬房 用薬
連文 (以上、第2回資料Ⅰより)

教界 航海権 新糸 攪み去る 不平均 (以上、第2回資料Ⅲより)

相対者 衛生者 消防式 葬祭場 年功者 拝賀式 (以上、第2回資料Ⅳより)

巨形 空中船 聴響器 (以上、第2回資料Ⅴより)

また、『日国大』に見出しはあるが用例のない語で、今回の実験で用例のえられたものには、つぎのようなものがある。

幾篇 痂皮 乾固 給水場 行政 口蹄疫 口内 国家経済 裁判所構
成法 商業界 貯蓄心 動物体 毒素 乳房 分泌液 癒合 (以上、
第1回より)

圧搾空気 運送費 海運業 偽薬 局内 合名会社 山村水郭 正価
正貨準備 製剤 大資本 代数学 着金 置郵 定期米 特約店 平作
平面幾何学 本誌 輸入超過 立体幾何学 和犬 和装 (以上、第
2回より)

これらのなかにも、『日国大』では用例を持ちながら、方針として用例をかかげることは省いたというものもあるかもしれない。

ところで、スカウトによって採集された用例は、総索引方式であればすべて採集されているはずのものであり、上に述べたような『日国大』との比較

は、それだけではスカウト方式の意義にかかわるものとはいえない。

いま、採集語の評価をスカウトという方式の問題と結びつけて考えるために、上掲の表16・17で得られた各一致度ごとの項目数が、1・2回の実験で得られた各一致ごとの項目総数のうちの何%をしめるかをみると、表18のようになる。

表 18 一致度と用例の有効性の関係

一致度 回	1	2	3	4	5	6	7	8	9	10	11	12	全体
1	13.3	17.0	24.8	28.4	42.9	35.3	48.5	52.2	66.7	90.9	42.9	—	24.5
2	10.8	12.3	23.4	24.0	26.3	43.0	42.3	65.5	50.0	64.4	60.7	93.8	25.7

* たとえば1回目の一致度1に対する13.3とは、1回目の実験で採集された一致度1の語数339(表1～4参照)に対して、表16で有効とされた語数45のしめるパーセントである。

この表から知られるように、『日国大』との比較からみて、前記のような基準で有効とされた用例は、一致度と密接な関連をもち、一致度の高いものほど、有効性が高いといえる。したがって、一致度が高いということは、その用例の有効性をはかる一つのめやすとなりうるわけであり、逆にいえば、採集者の訓練にあたっては、一致度を高める方向にみちびくということが有効な採集につながるといってよからう。

参考として、第2回の実験から、一致度の高かった語(一致度11および12)を列挙すれば、つぎのとおりである。

心育 神効 球算 端正優和 智育 重学 表語初学自修講義 (以上、資料Ⅰ)

盈虚 学政振張 学制調査会 軽重緩急 国語調査会 商量精査 秩録 処分 不可能的 (以上、資料Ⅱ)

相合同一致 哀請運動 外資輸入 株価回復策 官私線鉄道 教界 京仁鉄道 好望 私設鉄道 常套語 常務 真個 真主義 正貨準備 対清貿易 買取代価 (以上、資料Ⅲ)

下り者 月給取風 戸外運動 山村水郭 下た目 窓戸開放 追弔 鉄

道唱歌 東京化 年功者 流行り唄 牧童蚕婦 (以上, 資料Ⅳ)

空中飛行器 (以上, 資料Ⅴ)

ただ、今回の実験の範囲では、漢字3字・4字といった語をかたちのうえの長大きに着目して採集し、それが結果としては、『日国大』の立項の基準にはずれるものであるがゆえに、『日国大』に見出しのない語となったということもあろうかと思う。こうした長大な語形を含まない資料でも、一致度と、用例としての有効性とは結びつくものかいなかという問題は、用例の有効性をはかるあらたな基準の設定とともに、今後に残された課題である。

5. おわりに

今回の実験ははじめてのころみであり、参加者にスカウトという作業に対するとまどいがみられたのみならず、実験をおこなうがわにも、事前に問題の所在を十分明確にしえないというところがあった。結果の分析に際して、結論を出すに至らなかった問題があるのも、その原因の一部は、実験そのものに不備があったためである。だが、実験を進めるうえでのそうした困難は、単にふなれというわくをこえて、スカウト自体の流動性といったものを反映しているように思われる。そうしたむずかしさを事実として確認したことも、今回の実験の成果の一つといってよいだろう。その一方、個人間の異同の程度などを、数値として把握できたことは、スカウト方式に対する評価にめやすを与えるのに参考となる。

今回の限られた範囲内での実験から、あらためて検討を要する問題点も、いくつか生じてきている。その一つは資料の問題である。今回は、2回を通じて9種の資料をとりあげたが、そのほとんどが漢語の含有率の高い文語文である。口語体の小説などを資料にしたときに、スカウトはしやすくなるのかどうか、特に一致度は高まるかどうか。今回の実験でとりあげられた資料は、内容的には読みやすくないものが多かったかもしれないが、スカウトという点では、長大な漢語をめじるしにするということができたはずである¹⁰⁾。むしろ難解な語の含まれない資料から、適切なスカウトをおこなうことの方

が、いっそうむずかしいかもしれない。また、今回の実験では、1種の資料について2ページ程度の分量をあつかったのみである。実際のスカウトでは、一つの資料がかなりの分量を持つことになるであろう。そうしたことは、スカウトにどのように影響するものであろうか。

さらに、採集された用例を評価する判断の基準の問題がある。これは、うらがえせば、採集の基準そのものの問題である。今回は、多分に便宜的な方法として、『日本国語大辞典』との比較によったが、用例の価値は、その用例の使用者、あらわれる資料、文脈などを含めて、総合的に定められるべきであろう。また、どのような視点に立つかによっても、用例の有効性はかわってくるはずである。このように考えれば、用例の評価における絶対的な基準というものには存在しないというべきかもしれないが、限られた範囲内の用例収集をより有効なものとする方策として、どのような用例がよりのぞましいものであるかは、つねに反省されてしかるべきであろう。以上のような観点に立って、あらためて、実験試行がすみかさねられるべきである。

<注>

- 1) 本稿にいうところのスカウト方式による用例採集としてつとに著名なものとして、OED 編集の際の文献閲読者による協力があげられる。多くの有用な用例をえることができた反面、編者の意図を正しく理解してもらうために払われた苦心も大きなものであったようである。詳細は、OED の編者 J. R. Murray の伝記 “Caught in the web of words” (邦訳『ことばへの情熱』三省堂 1980) にみえる。わが国では、見坊豪紀氏による現代語を中心とした用例収集がよく知られているが、これは個人による統一的な問題意識にもとづく採集であり、一般の読者による用例採集とはことなるところがある。
- 2) 「国立国語研究所年報31」参照。なお、100語あたり50語を採集するということは、実際には、ことなり語をすべて採集するというのに近く、逆にいえば、ことなり語をすべて採集しなければ、個人間のことなりは消えないということであろう。
- 3) 採集量の点でいうと、ことなる語を網羅しようとするものから今回の実験でこころみたく1行あたり1個程度まで、3段階が考えられた。また、この試行で実際に採集された『坊つちゃん』の用語については、『表現の情報学』（講座

日本語の表現」6 筑摩書房 1983) に収められたくことばを調べる——辞書の使い方——に言及がある。

- 4) 『用例採集のための主要雑誌目録』は、原則として国立国会図書館所蔵の和雑誌から、①昭和25年以前創刊、②20年以上継続、の条件で選び出した1958誌の目録である。そこから、国語辞典編集準備調査会委員のかたがたにさらに選定をお願いし、10名中4名以上の推薦を受けたものとして、120誌が選ばれた。ちなみに、10名全員の推薦をえたのは、「太陽」のほか、「アララギ」「改造」「子供の科学」「帝国文学」「婦人公論」「文芸春秋」「ホトトギス」の計8誌であった。
- 5) 採集量を行数との対比で定めるのは、1行あたりの字数の違いなどを考えれば、多分に便宜的なものといわなくてはならない。なお、1回目の実験の資料についていうと、Ⅱが約430文節、Ⅰ・Ⅲ・Ⅳが約770~780文節である。
- 6) このなかには、かなづかいや字体に着目して採集したために、採集数から除外されたものもあるが、それは少数の参加者の場合のみである。
- 7) 実験後の参加者の感想でも、採集数を1行あたり1個程度にそろえることのむずかしさが指摘された。とくに、いったん採集したあとで、追加して採集するということが困難なようである。
- 8) この比較には、スカウトに対するふなれ、とくに一方が第1回目であるという特殊性が含まれている。スカウトになれたのちに、同資料で2回採集をおこなえば、またちがった結果がえられるかもしれない。
- 9) こうした語形が見出しには立てられず、かといってその構成要素にあたる項目の用例としてもあげられないというのは、従来の国語辞書の不備の一つといってよからう。
- 10) 今回の実験では長い単位で採集してもらったが、実際の採集では、長短2種類の単位を併用することが考えられている。

<付記>

本稿のもととなった実験の企画と実施は、飛田良文(国語辞典編集準備室主任)・高梨信博(書記)・見坊豪紀・林大(以上、調査員)が担当した。また、本稿は高梨がまとめたが、稿をなす過程で、上記のかたがたにお教えを受けたところが多い。また、実験に被験者として参加したのは、国語辞典編集準備室関係では、上記の4名のほか、滝本典子・中田恵美子(以上、調査員)、伊藤真一郎・妹尾和子・平沢啓・吉田ひろみ(以上、アルバイト)、外部からお願いしたのは、荒井早苗・大沢とく・大道八重子・小口三佳子・末広美代子・前田富士子のみなさんである。しるしてあつくお礼申しあげる。