

国立国語研究所学術情報リポジトリ

国定読本における類義語の現われ方：
「うつくしい」と「きれい」

メタデータ	言語: Japanese 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): Dictionary, Corpus, Concordance, Synonym, Kokutei Tokuhon 作成者: 加藤, 安彦, KATŌ, Yasuhiko メールアドレス: 所属:
URL	https://doi.org/10.15084/00001144

国定読本における類義語の現われ方

— 「うつくしい」と「きれい」 —

加 藤 安 彦

KATŌ Yasuhiko: The Appearance of Synonymous Words in
“Kokutei Tokuhon” —*utsukushii* and *kirei*

要旨：辞典編集という観点からすると、集められたデータの中からそれぞれの見出し語に対し、用法や意味の異なる例を多く載せることが望ましいと考えられる。しかし、そうした辞書データを作成する場合に作業者が何万とある用例それぞれについて用法や意味の判断を施していくことはあまり効率的とはいえない。意味が近いと思われる語の用例を集めて作業を行えば、意味や用法の同じものを束ねることが容易となろうし、その中から代表的な例を挙げるのであれば作業効率も向上すると思われる。本稿では、『分類語彙表』を利用して「うつくしい」と「きれい」という類義語を比較し、各々どのような点が共通しているかを考察する。また、意味の近さを判断するための客観的尺度の有効性についても述べる。

キーワード：辞書，コーパス，コンコーダンス，類義語，国定読本

Abstract: From the standpoint of editing a dictionary, it is desirable to list examples of different usage or meaning from collected linguistic data. But to form a judgment on whether or not to use each one of thousands of examples is not effective way to make dictionary data. Working on gathered examples of synonymous words makes it easy to classify them by meaning or usage, and promotes efficiency.

In this paper, I refer to “Bunrui-goi-hyo” to clarify the common features between the synonymous words “*utukusii*” and “*kirei*” which appear in Kokutei Tokuhon text data, and also to the effectiveness of the thesaurus as an objective criterion for synonymity.

Key Words: Dictionary, Corpus, Concordance, Synonym, Kokutei Tokuhon

1. はじめに

「国定読本用語総覧」は国語辞典編集室がその目標として掲げている国語辞典編集のための準備資料である。国定読本は第一期から第六期までの六期に分かれるが、第五期国定読本の「あ～つ」が「国定読本用語総覧 第8巻」として1993年8月に刊行された。

国語辞典編集のための準備資料は国定読本以外にも文学作品、雑誌などジャンルも多岐にわたる。採集されるのは語のみではなく、その周辺、いわばその語が使用されている環境である文の前後の部分も採集される。こうして採集したデータが蓄積されて十分に大きな数となった時、用例採集と並行して国語辞典編集の基本的な作業が始まることになるが、採集した何十万、何百万とある用例すべてを辞典の用例として採用することは難しい。国語辞典ということを見ると研究用データと事情がちがいで、できるだけ用法、意味の異なる用例を多く採用することが望ましい。しかしながらその辞典用データを作成する場合、ただランダムに集められた用例に対し、作業者が一々必要となる情報について判断を施していくという方法では、あまり効率的とはいえない。そこで、ある程度意味が近いとの予測がなされてまとめられている用例の集合に対して作業を行えば、意味や用法の同じものを束ねることが容易となろうし、その中から代表的な例を挙げていくのであれば作業効率も向上するだろうと思われる。ただし、意味が近そうなものを集めるとしても、あらかじめ出現するであろう意味を想定してデータにあたり、分類していくという方法は避けた方がよいと考える。別扱いして考えることが望ましいかもしれないもまで想定した何れかの意味に押し込んでしまうおそれがないとも限らないからである。本稿で、「うつくしい」と「きれい」という二語の用例について個々の意味を云々するところから始めなかったのはこの理由による。各々の語の意味をあらかじめ想定するのではなく、それぞれの用例全体をまず一つの「かたまり」として捉えることから始め、語の内側にある情報によってではなく、分類語彙表などいわば外側の客観的尺度によって、用例を束ねてみたかったのである。ここではこの二語がどのように現わ

れ、どのように使われているかについても報告するが、作業者が不特定かつ複数にわたるような辞書データ作成の今後の方針に関わるなんらかの手掛かりが得られれば、と考えている。

2. データについて

2.1. 対象

対象としたのは、第一期から第五期の国定読本に出現した「うつくしい」と「きれい」の用例すべてであるが、「うつくしい」には「うつくし」の用例も含めてある（「うつくし・うつくしい」は、特に断わらない限り、以下「うつくしい」と記す）。五期を通じての出現総数は「うつくしい」が272例、「きれい」が155例ある。

「国定読本用語総覧」において「うつくしい」は「形容詞」、「きれい」は「形状詞」である。本稿で「きれい」として扱っているのは、終助詞「ね」を伴う一例以外、形状詞「きれい」に助動詞「だ」あるいは「です」を介した用例である。

2.2. 語への付加情報について

テキスト内に出現した語すべてが網羅される採集方式が、全数式用例採集法であるが、そこで得られた自立語に分類語彙表の番号を機械的に割り振ることは可能であり、また、修飾－被修飾といった係り受け関係も構文解析システムがあれば結果として得ることもできる。ここでは、こうした操作をすでに経ているデータと仮定して「うつくしい」と「きれい」の二語の例を扱うことにした。

2.2.1. 用法の情報

上述の仮定された構文解析による係り受けの解析結果は、以下のように用法を修飾用法と述語用法の二つに大別し、修飾用法はさらに体言修飾と用言修飾の二つに分けることにする。

【用法】

○修飾用法

体言修飾

用言修飾

○述語用法

ここでいう修飾用法とは、

「うつくしい夕日」（＝体言修飾）

「きれいな模様」（＝体言修飾）

「カラスウリがうつくしく色づいた」（＝用言修飾）

「いろんな色をきれいに使う」（＝用言修飾）

といった体言または用言を修飾する用法であり、述語用法は、

「夕日がうつくしい」

「色がきれいだ」

といった主語を持つ用法を指す。

出現用例すべてに対して、修飾用法であるか、述語用法であるか、修飾用法であれば、体言修飾であるか、用言修飾であるかという判断は今回すべて人手によって行った。以下に述べる分類語彙表の番号付与も同様に人手によった。

2.2.2. 分類語彙表の情報

前述の修飾用法において、体言修飾ならば、被修飾体言の属する意味カテゴリー、用言修飾ならば、被修飾用言の属する意味カテゴリー、述語用法においては、主語となっている語の属する意味カテゴリーのそれぞれがどのよ

うな分類に該当するかを国立国語研究所資料集6「分類語彙表」(第28版)を用いて検討し、その分類番号の付与を行った。

述語用法では、必ずしも主語が存在するというわけではなく、「ああ、きれいだ。」といった文(これを本稿の中では無主語文と呼んでおく)もあり、その場合には、なにを指しているのか、文脈から判断して本来主語となるべき語の分類番号を付与してある。また、被修飾語あるいは主語となる語が接続助詞などで結ばれて複数存在する場合には、それを各々独立させて考えてあるため、データ上では「うつくしい」および「きれい」の出現数を被修飾語、主語となる語の分類語彙表の分類番号別にまとめた総数が上回っている場合がある。

「うつくしい」と「きれい」とが類義であることは、分類語彙表において「3. 502」に共に分類されていること、また、国定読本の中においても、同一対象に対して同時に用いられている例があること(第四期第五巻「うつくしい尾を、扇のやうにひろげました。『まあ、きれいだこと。<略>』)から、類義語としてよいだろう。ただし、両者の間には一方と意味が異なる、あるいはある用法が一方にあって一方にないという場合がある。例えば、「皿をきれいに洗う」、「出された食事をきれいにたいらげる」という時には、「汚れ・ごみなどのない状態」、「なにも残っていない状態」を「きれい」というわけであり、「うつくしい」ではそうした意味を担えない。逆に「うつくしい兄弟愛」が意味する「うつくしさ」を「きれい」が担うことはできない。こうした点を考慮する必要があるが、個々の例にあたって意味を一々判断する方式は将来的な方針とずれるため、今回はそうした判断を用法の情報による係り受け関係と、その係り受け関係にある語の分類語彙表上の番号によってどこまで判断がつけられるのかを検討することにした。

2.2.3. 層別情報

上述の情報以外に、個々の例について考慮したのは以下に挙げる層別情報である。

【層別情報】

会話文

韻文

手紙文

上記以外=地の文

層別情報は、「国定読本用語総覧」に付けられている情報で、このほかに文語文、候文といった情報があるが、今回は文語も含めて扱っているので文語文の情報は考慮せず、また、候文は用例の中に存在しなかった。ここで地の文としているのは「国定読本用語総覧」上に層別情報のないものをいう。

3. 「うつくしい」と「きれい」の比較

国定読本の第一期から第五期までのデータは、教育のための読本という共通部分を持ちながらもそれぞれは独立した文の集合体で、いわば独立したコーパスであるといえる。その個々の中において、また全体を通じて「うつくしい」と「きれい」とがどのような使い分け、役割分担があるのかをみることにする。

3.1. 総出現数

「うつくしい」および「きれい」の出現数は、表1の通りである。各期の総延べ語数も挙げておく。

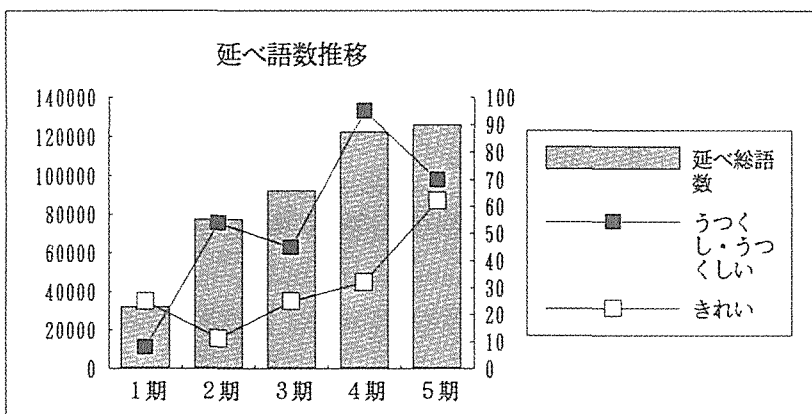
	1期	2期	3期	4期	5期
総延べ語数	32362	77358	92010	122429	126033
うつくしい	8	54	45	95	70
きれい	25	11	25	32	62

表 1

これらの出現数の推移を「国定読本用語総覧」各期の総延べ語数の推移とともに表わしたのがグラフ1である。左の縦軸は総延べ語数に対応する語数

であり、右の縦軸は「うつくしい」と「きれい」の双方の延べ語数に対応する語数である。

これによると、「うつくしい」と「きれい」は若干の上下はありながらも総延べ語数の第一期から第五期への増加に伴って変化をみせていると見てよい。いうなれば国定読本での「普通に使われている語」である。第六期までの資料を作成した段階でこのボトムアップな視点からの「なにか国定読本における基本的な語であるか」という判断があってもよいと思われる。



グラフ 1

3.2. 各期各学年別・用法別出現数

「うつくしい」、「きれい」の各期における各学年別、用法別の出現数をまとめたものが、表2、表3である。

修飾用法出現総数と述語用法出現総数の和が学年別出現数である。また、体言修飾と用言修飾の和が修飾用法出現総数である。

※第二期以降は第一学年から第六学年までで教科書は十二巻あるが、第一期については第四学年までの八巻だけであるので、グラフなどにすると若干粗い変化を見せる場合がある。

うつくし・うつくしい	1期							2期						
	一年	二年	三年	四年	五年	六年	総数	一年	二年	三年	四年	五年	六年	総数
修飾用法出現総数	0	2	0	2			4	3	7	7	9	11	3	40
体言修飾出現数	0	2	0	2			4	3	7	7	9	10	2	38
用言修飾出現数	0	0	0	0			0	0	0	0	0	1	1	2
述語用法出現総数	0	2	2	0			4	1	3	6	2	2	0	14
(内無主語文数)	0	0	0	0			0	1	2	1	1	0	0	5
学年別出現数	0	4	2	2			8	4	10	13	11	13	3	54
	3期							4期						
	一年	二年	三年	四年	五年	六年	総数	一年	二年	三年	四年	五年	六年	総数
修飾用法出現総数	0	2	5	6	12	7	32	0	6	7	21	23	30	87
体言修飾出現数	0	2	4	6	12	7	31	0	6	6	15	16	20	63
用言修飾出現数	0	0	1	0	0	0	1	0	0	1	6	7	10	24
述語用法出現総数	0	0	3	3	5	2	13	0	0	0	1	2	5	8
(内無主語文数)	0	0	0	0	1	0	1	0	0	0	0	0	1	1
学年別出現数	0	2	8	9	17	9	45	0	6	7	22	25	35	95
	5期													
	一年	二年	三年	四年	五年	六年	総数							
修飾用法出現総数	1	5	7	14	10	18	55							
体言修飾出現数	1	3	6	8	7	10	35							
用言修飾出現数	0	2	1	6	3	8	20							
述語用法出現総数	0	5	0	2	3	5	15							
(内無主語文数)	0	1	0	0	2	0	3							
学年別出現数	1	10	7	16	13	23	70							

表 2

表 2, 表 3 の数値に基づいて, それぞれ「うつくしい」, 「きれい」の期毎の出現総数に対する各学年毎の出現数比率を求めたのが表 4, 表 5 であり, それを折れ線グラフにしたものがグラフ 2, グラフ 3 である。Y 軸はパーセンテージ, X 軸は学年を表わす。

第一期から第五期までの全体的な傾向を眺めてみると「うつくしい」は低学年から高学年に向けて増加をみせており, 「きれい」は逆に二年, 三年を中心に高学年に向けて減少をみせている。「うつくしい」, 「きれい」が個別の用例においてどのような意味で用いられているか調べる必要があり, 短絡的に結論づけるわけにはいかないが, 同様の事象に対して低学年では「きれい」を, 高学年では「うつくしい」を用いて表現するといった両者の出現上の「すみわけ」も可能性としてあるのではないかと思われる。

きれい	1期							2期						
	一年	二年	三年	四年	五年	六年	総数	一年	二年	三年	四年	五年	六年	総数
修飾用法出現総数	2	7	1	5			15	0	2	2	2	1	1	8
体言修飾出現数	1	4	1	3			9	0	2	1	0	0	0	3
用言修飾出現数	1	3	0	2			6	0	0	1	2	1	1	5
述語用法出現総数	0	8	2	0			10	1	2	0	0	0	0	3
(内無主語文数)	0	5	1	0			6	1	1	0	0	0	0	2
学年別出現数	2	15	3	5			25	1	4	2	2	1	1	11
	3期							4期						
	一年	二年	三年	四年	五年	六年	総数	一年	二年	三年	四年	五年	六年	総数
修飾用法出現総数	1	5	6	1	4	1	18	1	6	8	2	1	2	20
体言修飾出現数	1	2	5	1	3	0	12	0	5	6	2	0	0	13
用言修飾出現数	0	3	1	0	1	1	6	1	1	2	0	1	2	7
述語用法出現総数	0	2	1	0	3	1	7	2	2	5	1	0	2	12
(内無主語文数)	0	2	0	0	1	0	3	2	1	2	0	0	2	7
学年別出現数	1	7	7	1	7	2	25	3	8	13	3	1	4	32
	5期													
	一年	二年	三年	四年	五年	六年	総数							
修飾用法出現総数	5	11	11	11	3	3	44							
体言修飾出現数	2	7	7	9	0	1	26							
用言修飾出現数	3	4	4	2	3	2	18							
述語用法出現総数	2	6	4	4	2	0	18							
(内無主語文数)	2	1	2	2	2	0	9							
学年別出現数	7	17	15	15	5	3	62							

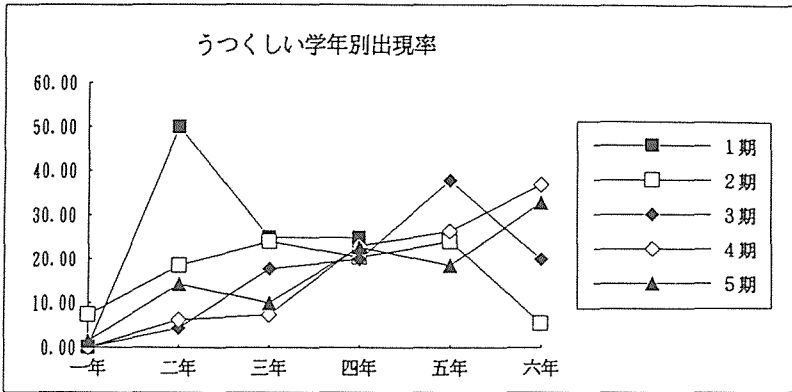
表 3

うつくしい学年別出現率	一年	二年	三年	四年	五年	六年
1期	0.00	50.00	25.00	25.00		
2期	7.41	18.52	24.07	20.37	24.07	5.56
3期	0.00	4.44	17.78	20.00	37.78	20.00
4期	0.00	6.32	7.37	23.16	26.32	36.84
5期	1.43	14.29	10.00	22.86	18.57	32.86

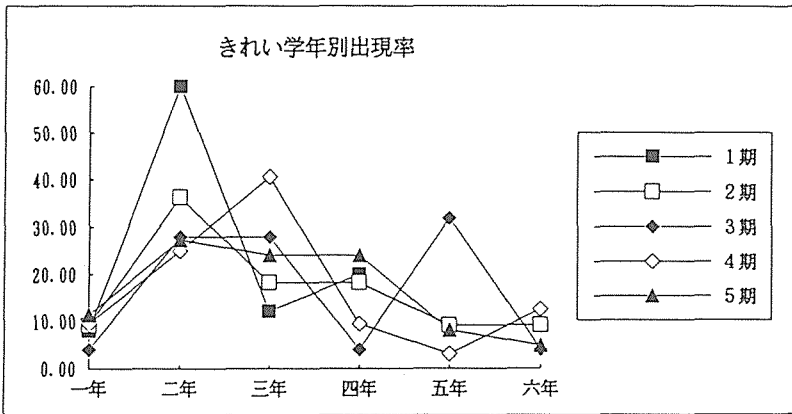
表 4

きれい学年別出現率	一年	二年	三年	四年	五年	六年
1期	8.00	60.00	12.00	20.00		
2期	9.09	36.36	18.18	18.18	9.09	9.09
3期	4.00	28.00	28.00	4.00	32.00	4.00
4期	9.38	25.00	40.63	9.38	3.13	12.50
5期	11.29	27.42	24.19	24.19	8.06	4.84

表 5

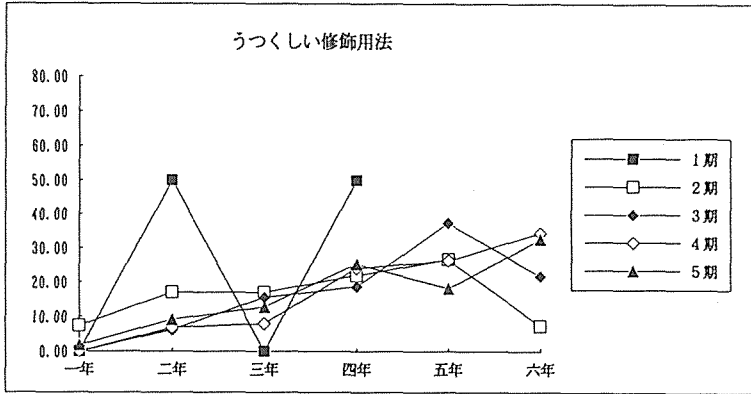


グラフ 2

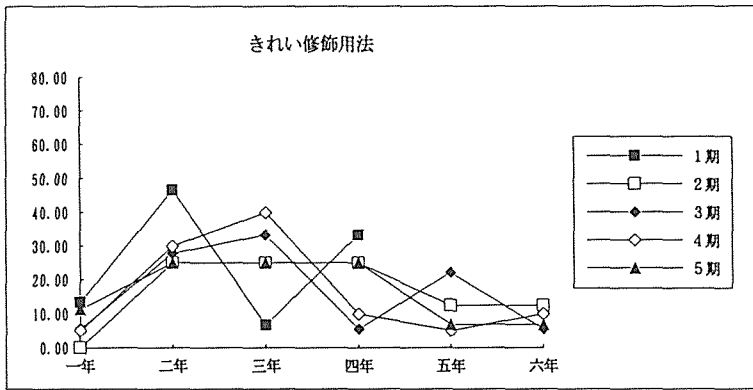


グラフ 3

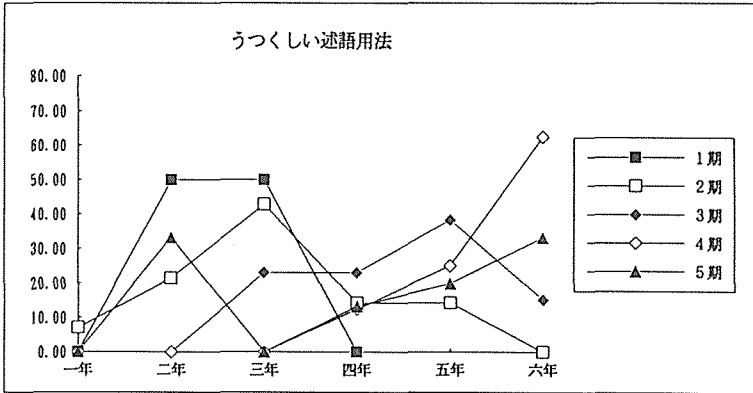
次に、表 2、表 3 に基づいて、期毎の用法別総数に対する各学年の出現数の比率を求めたのがグラフ 4 からグラフ 7 である。



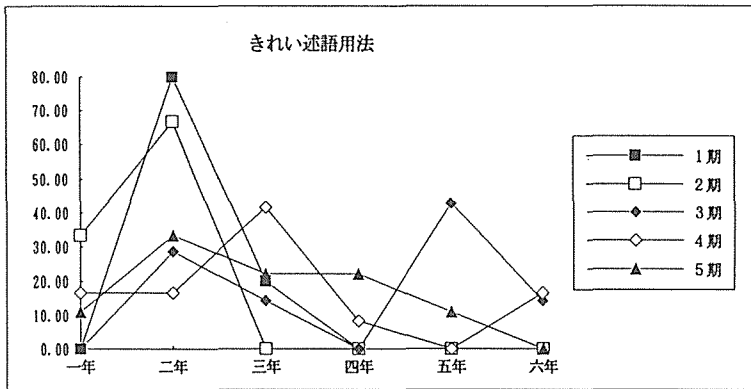
グラフ 4



グラフ 5



グラフ 6



グラフ 7

修飾用法については、「うつくしい」、「きれい」ともに変化のしかたが学年別出現率の変化を示したグラフと同様である。

述語用法は、修飾用法に比べると、期毎の変化のしかたにばらつきがあり、用いられ方が一定ではないといえる。「うつくしい」第四期のグラフのように第四学年になるまでまったく出現しないものもあれば、「きれい」第二期のように第三学年以降まったく出現しないものもある。出現しているものの意味を個々に調べてみたが、例えば「きれい」が「汚れ・ごみなどのない状態」であれば低学年、「なにも残っていない状態」ならば高学年というよう

なはっきりとした説明可能な傾向をみることはできなかった。

3.3. 層別情報

各学年別出現数の総数に対する層別情報それぞれの出現数、およびその比率を求めたのが表6、表7であり、そのグラフ化したものが以下のグラフ8、グラフ9である。

層別の出現変化は、

「うつくしい」……第一期では偏りがあまりなく、「きれい」と同様、各層に分散して使われているが、第二期以降はもっぱら地の文に偏って出現している。

「きれい」……地の文での出現率は高いが、第二期以降減少傾向にある。

偏りが「うつくしい」よりは少なく、会話文での出現率が特徴的である。

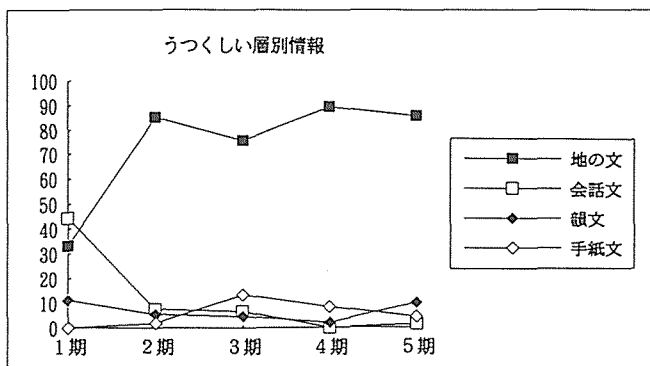
という傾向をみることができ、ここでは「うつくしい」が文章語的な性格を第二期以降の変化で見せており、「きれい」はそれに対して口語的な部分を担っていることが伺える。(同じ日の出に対する表現として、第二期においては「ウツクシイデハアリマセンカ。」と地の文として書かれているものが、第三期では「ア、日が出ハジメタ。キレイダ。」と会話文として書かれているような例にそれが見られる。)

うつくしい		1期		2期		3期		4期		5期	
		総数	%	総数	%	総数	%	総数	%	総数	%
学年別出現数		9		54		45		95		70	
層別情報	地の文	3	33.33	46	85.19	34	75.56	85	89.47	60	85.71
	会話文	4	44.44	4	7.41	3	6.67	0	0.00	1	1.43
	韻文	1	11.11	3	5.56	2	4.44	2	2.11	7	10.00
	手紙文	0	0.00	1	1.85	6	13.33	8	8.42	3	4.29

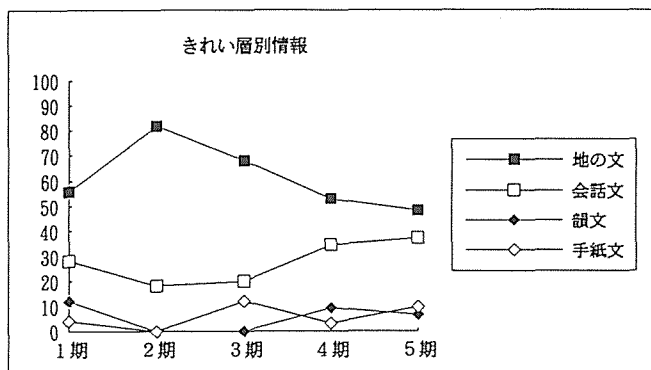
表6

きれい		1期		2期		3期		4期		5期	
		総数	%	総数	%	総数	%	総数	%	総数	%
学年別出現数		25		11		25		32		62	
層別情報	地の文	14	56.00	9	81.82	17	68.00	17	53.13	30	48.39
	会話文	7	28.00	2	18.18	5	20.00	11	34.38	23	37.10
	韻文	3	12.00	0	0.00	0	0.00	3	9.38	4	6.45
	手紙文	1	4.00	0	0.00	3	12.00	1	3.13	6	9.68

表 7



グラフ 8



グラフ 9

3.4. 分類語彙表番号情報

分類語彙表番号は、以下のものに割り振った。

修飾用法

体言修飾……………被修飾体言

用言修飾……………被修飾用言

述語用法

述語用法文の主語となる語

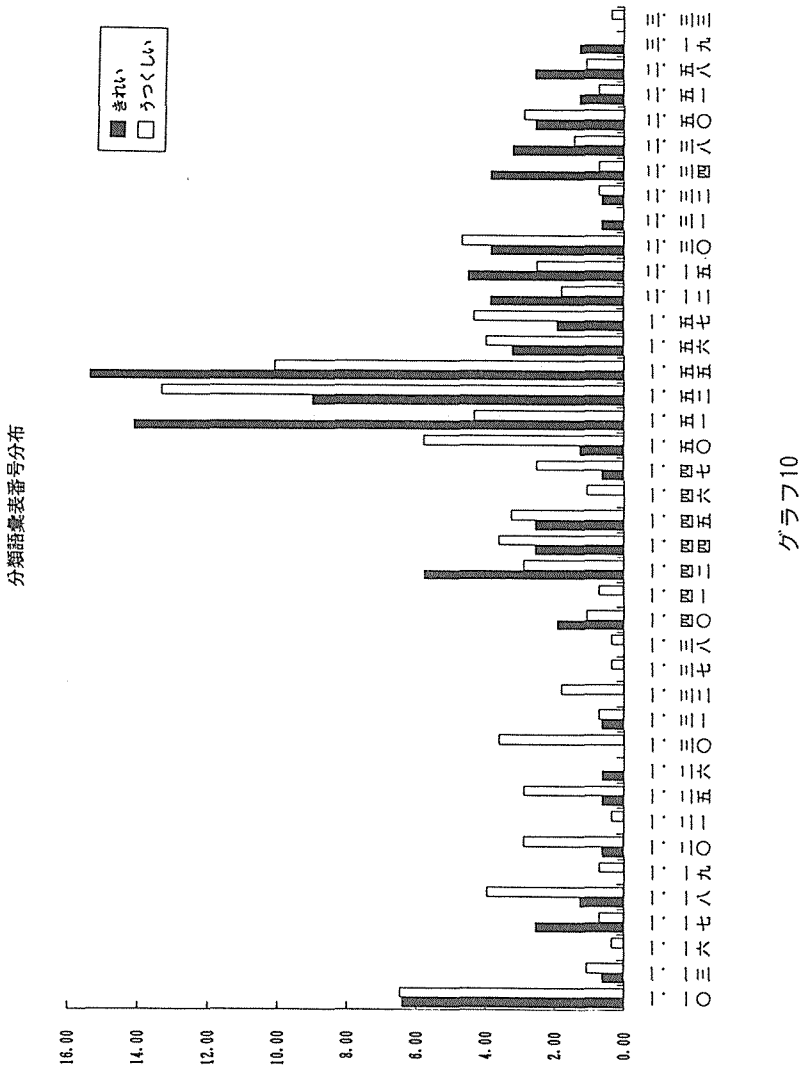
無主語文の場合、文脈から主語と考えられる語

各々の語に振られた番号は、六桁の数字で、「分類語彙表」のインデックスによって得られるものである。グラフを作成するのに六桁までの番号を用いると細かすぎるため、まず、三桁までの分類番号を用いて比較してみる。表8は、「うつくしい」および「きれい」の被修飾語あるいは主語となる語、主語と考えられる語の分類カテゴリーがそれぞれの総数に対してどの程度の割合を占めているかという比率を表わしたものである。

	一、一〇	一、一三	一、一六	一、一七	一、一八	一、一九	一、二〇	一、二一	一、二五
うつくしい	6.47	1.08	0.36	0.72	3.96	0.72	2.88	0.36	2.88
きれい	6.41	0.64	0	2.56	1.28	0	0.64	0	0.64
	一、二六	一、三〇	一、三一	一、三二	一、三七	一、三八	一、四〇	一、四一	一、四二
うつくしい	0	3.6	0.72	1.8	0.36	0.36	1.08	0.72	2.88
きれい	0.64	0	0.64	0	0	0	1.92	0	5.77
	一、四四	一、四五	一、四六	一、四七	一、五〇	一、五一	一、五二	一、五五	一、五六
うつくしい	3.6	3.24	1.08	2.52	5.76	4.32	13.31	10.07	3.96
きれい	2.56	2.56	0	0.64	1.28	14.1	8.97	15.38	3.21
	一、五七	二、一二	二、一五	二、三〇	二、三一	二、三二	二、三四	二、三八	二、五〇
うつくしい	4.32	1.8	2.52	4.68	0	0.72	0.72	1.44	2.88
きれい	1.92	3.85	4.49	3.85	0.64	0.64	3.85	3.21	2.56
	二、五一	二、五八	三、一九	三、三三					
うつくしい	0.72	1.08	0	0.36					
きれい	1.28	2.56	1.28	0					

表 8

表8をグラフにしたものが、グラフ10である。



グラフ10で、「うつくしい」の数値が高いものが26カテゴリー、「きれい」の数値が高いものが14カテゴリーある。

この「きれい」の数値が高いとされる意味カテゴリーと、先に述べた「きれい」固有の「汚れ・ごみなどのない状態」、「なにも残っていない状態」といった「うつくしい」の担えない意味で用いられている用例との関係のみをみることにする。

個々の用例にあたり、第一期から第五期までに出現した「きれい」固有と考えてよい用例が40例ほどあり、それらの被修飾語、主語となる語の分類語彙表番号を集計してみた。結果は以下の通りであるが、これに「きれい」の数値が「うつくしい」の数値を上回った意味カテゴリーを重ねて考えると、「うつくしい」の数値を上回った14カテゴリーの内、9カテゴリーが「きれい」固有の意味を持つ用例における被修飾語、主語となる語の意味カテゴリーとして出現したものであった。しかし、「きれい」の数値が上回

ている他の5カテゴリーがここではみられず、また、出現数としては少ないが、グラフで「うつくしい」の数値が上回っているカテゴリーも4つ含まれていた。

出現数	カテゴリー
12	1. 51
5	2. 34
4	2. 50
4	2. 12
3	2. 15
2	1. 17
2	1. 52
2	2. 30
2	2. 38
2	2. 51
1	1. 40
1	1. 42
1	1. 57
1	2. 32

他の分析要素が必要となるが、おおまかな判断をくだす基準として分類語彙表番号の付与が有効であることは確かなようである。

複数の作業員によってデータ作成を行う場合に問題となるのは、個々の作業の判断のゆれであるが、それをできるだけ吸収して作業結果の均一化を図っていかなければならない。分類語彙表はその点、機械的な番号付与も可

能で、作業のゆれを少なくするための補助的な情報として有効といえるだろう。

4. おわりに

最後に全数式用例採集法における全体の延べ語数と各語の延べ出現数の関係を用いて今後検討してみたい点について述べておく。

「国定読本」というテキストにおける文の集合は、それ自体「生コーパス」(raw corpus)であるが、「国定読本用語総覧」という、ある単位に基づいて語の認定が行われ、品詞情報も与えられているデータがその背景としてあるので、「国定読本用語総覧」を用いた「国定読本」テキストの形態素解析を行えば、容易に「タグ付きコーパス」(tagged corpus)に変換できる。

ここにコーパス i とコーパス j がある場合、それぞれが異なりで n 個と m 個の語により構成されている。これらを集合論的な記法によって以下のように表わすことができる。

$$\Sigma i = \{a_1, a_2, \dots, a_n\}$$

$$\Sigma j = \{b_1, b_2, \dots, b_m\}$$

ここで、コーパス i における語 W の延べ出現総数を $n_i(W)$ で表すことにする。

このとき、コーパス i における総語数を

$$N(i) = \sum_{k=1}^n n_i(a_k)$$

と表わし、またコーパス j における総語数を

$$N(j) = \sum_{k=1}^m n_j(b_k)$$

と表わすことができる。

ここで、

$$N(i) = N(j) \times \theta 1$$

という式によって、コーパス i およびコーパス j における総語数の関係を表すことができる。 $\theta 1$ は、コーパス i およびコーパス j における総語数推移の比率である。

コーパス i における Σi 、コーパス j における Σj にそれぞれ同一の語 α が含まれているとすると、

$$n_i(\alpha) = n_j(\alpha) \times \theta 2$$

という式で両者間の関係を表わすことができる。この $\theta 1$ および $\theta 2$ について、ほぼ同傾向の推移比率とみなされる場合、語 α はコーパス i、コーパス j を統合したコーパスにおける基本語としてよいのではないか、あるいはそうした視点での基本語を考えられないだろうか。この視点は、語 α の使用コーパスのジャンルにも、コーパス内での役割にも依存しないことになる。ある時間的な区切りを設け、その期間内におけるコーパスで「ふつうに使用されている語」ということである。

国定読本の第一期から第六期までのデータが揃った時点で、各期の総延べ語数推移の比率と出現数推移の比率に正の相関がある語を「国定読本」コーパスの基本語、ふつうに使用されている語とし、さらには、全数式用例採集によって国定読本に付加されていく他ジャンルのデータとの間においても同様の変化をみせるものを抽出していく。これを現代第二期（1901～1950）という期間における傾向把握に関わらせていってみたい。もちろんこれは、コーパスという有限個の語からなる閉じた集合の中での話で、「教育基本語彙」といったものとはもとより意味合いがちがうものである。さらにそれを他期

間との比較にもつなげてみたいが、全語について調査してみない限り「アイデア」の域を出ない話であり、それらを基本的なものといってよいかどうかははっきりしないものである。今後十分な数の例を集め、検証していく必要がある。

参考文献

国立国語研究所：国立国語研究所国語辞典編集資料 1～8 「国定読本用語総覧 第1巻～第8巻」

国立国語研究所：国立国語研究所資料集 6「分類語彙表」第28版，秀英出版(1990)

Nirenburg, S. : Lexicon Acquisition for NLP :
A Consumer Report, CMU (1990)

Francis, W. N. and, Kucera, H. : Standard Corpus of Present-Day Edited
American English, for use with Digital
Computers, Brown University (1964)

Svartvik, J. and Wekker, H. : Topics in English Linguistics 3, Eng-
lish Computer Corpora; Selected Pa-
pers and Research Guide, Mouton de
Gruyter (1991)