高校教科書の語彙調査

中野　洋，土屋　信一，鷲岡　昭夫
（国立国語研究所）

0．はじめに

昭和49年に着手した高等学校の社会科・理科を対象とした語彙調査の作業が終了し、昭和58年2月に語彙表を公刊した。この調査は昭和41年の新聞三紙を対象とした語彙調査に続くものであるが、より一層計算機処理を多く採り入れ、語彙表までも高速漢字プリンターで打ち出したものをそのまま版下にしている。今後も各種の語彙表・集計表を作成し、分析・記述を進めていく予定である。

本報告では、調査の概要と語彙の量的構造について報告する。

1．調査の概要

（1）調査の目的　現代日本語の用語用字の実態を明らかにするために，国立国語研究所では，これまで，新聞，婦人雑誌，総合雑誌，雑誌九十種，新聞三紙を対象として，語彙調査を重ねてきた。とくに，昭和41年の新聞三紙を対象とした調査は，電子計算機を使用した，最初の大規模な調査であった。これらの調査のあとを受けて，国民が一般教養として、各分野の専門知識を身につける時に必要と思われる語彙の実態を明らかにすることを目的として，この高校教科書の語彙調査は企画された。高等学校進学率の増加に伴い，現今では、高等学校教育は，国民大多数の基本的な教養の場となっている。また，大学教育は、この高校教育の基盤に立って進められるものであり，とくに高校の理科と社会は，大学における専門教育の基礎となっていると考えることが出来る。われわれが高校の理科・社会科の全教科を対象とした語彙調査を企画したのは，以上のような理由からである。

なお，本調査は，　出現した語彙・表記・表現の実態を把握・分析することを目標としている。とくに，知識体系の記述を分析するために，この調査では，従来のようなサンプリング法によらず，対象とする文章を限定したのち，その全文を入力するという方法を採った。このやり方では，調査対象の幅をせばめてしまい，高校教科書の全体像を記述するという点では、やや不十分な結果しか得られない面もあるが，文章解析などこれまでの語彙調査では出来なかった数々の分析と記述を可能にしたと言える。

（2）調査対象　昭和49年度に使用されていた，以下の教科書の本文（巻末索引・年表，図表，脚注，人名・地名の上下についているアルファベット表記・生没年，下付きルビなどを除く）である。

（教科名）	（教科書名）	（著者名）	（出版社名）	（発行年月日）	
物理Ⅰ	標準高等物理	大塚明郎ほか	講談社	昭49．1．30	
化学Ⅰ	化学Ⅰ	柴田雄次ほか	大日本図書	〃49．2．5	
生物Ⅰ	生物Ⅰ	石田寿老ほか	清水書院	〃50．2．15	三版
地学Ⅰ	地学Ⅰ	湊　正雄ほか	実教出版	〃49．1．25	
倫理社会	倫理・社会	中村　元ほか	東京書籍	〃49．2．10	
政治経済	政治経済 新訂版	辻　清明ほか	自由書房	〃49．2．5	

日本史	詳説日本史(再訂版)	宝月圭吾ほか	山川出版	昭49．3．5
世界史	三省堂新世界史	土井正興ほか	三省堂	〃49．3．30再版
地理B	高校新地理B	青野壽郎ほか	二宮書店	〃49．1．20

（3）調査単位　長い単位と短い単位の2種類の調査単位を用いた。長い単位は，文の構成にあずかる要素（いわゆる文節）にもとづく単位で，word の頭文字をとってW単位と名付けた。短い単位は，語の構成にあずかる要素（いわゆる最小単位）にもとづく単位で，morpheme の頭文字をとってM単位と名付けた。

以下はW単位（／）とM単位（／および：）によって実際の文章を切った例である。単位切り規則については文献を参照されたい。

／ピルビン：酸／は／脱：炭酸：さ：れ：て／，／活性：化：さ：れ：た／酢酸／に／な：た／のち／，／まず／オキサロ：酢酸／と／反応：し：て／クエン：酸／となり／，／脱：水酸：酵素／の／はたらき／で／水素／を／失い／，／脱：炭酸：酵素／の／はたらき／で／二：酸化：炭素／を／失う／反応／を／くり：かえす／うち／，／図：4：2／の／よう／に／オキサロ：酢酸／に／もどる／。／

（4）同じ語か異なる語かの判別　同語異語判別とは，一語の範囲を定め，それに属する語に同じ見出し語を付することである。具体的には，異形同語に同じ見出しを付することと，同形（同表記）異語を分離しそれぞれに異なった見出しを付することの二作業に分れる。例えば，動詞「とる」は次のように21の異形態を持ち，一つの語と認められる。

「とる」 760	物理	化学	生物	地学	倫社	政経	日本史	世界史	地理
採っ						7			
執っ					1				
取っ		7			1	4			
とっ	25	16	20	7	51	37	34	30	11
撮っ									1
採ら						2			
取ら					1				
とら	1		1		4	7	5	2	2
採り						8			
取り	15	8		4		8	9		1
とり	25	42	60	10	44	3	39	38	10
採る						3			
取る		6			1	6			
とる	31	13	14	12	8	8	20	7	4
取れ		1							
とれ	11			1		2			
取れる					1				
とれる	1		1	1		1			1
とる					2		1		
どっ							1		
どる							1		

一つの見出しに集まった異形態の数で語を分類すると次のようになる。

異形態数	見出し数	%
1	14311	91.37
2	732	4.67
3	248	1.58
4	146	0.93
5	66	0.42
6	54	0.34
7	36	0.23
8	24	0.15
9	18	0.11
10	6	0.04
11～	21	0.13
計	15662	

11以上の異形態を持つ見出しは次の語である。

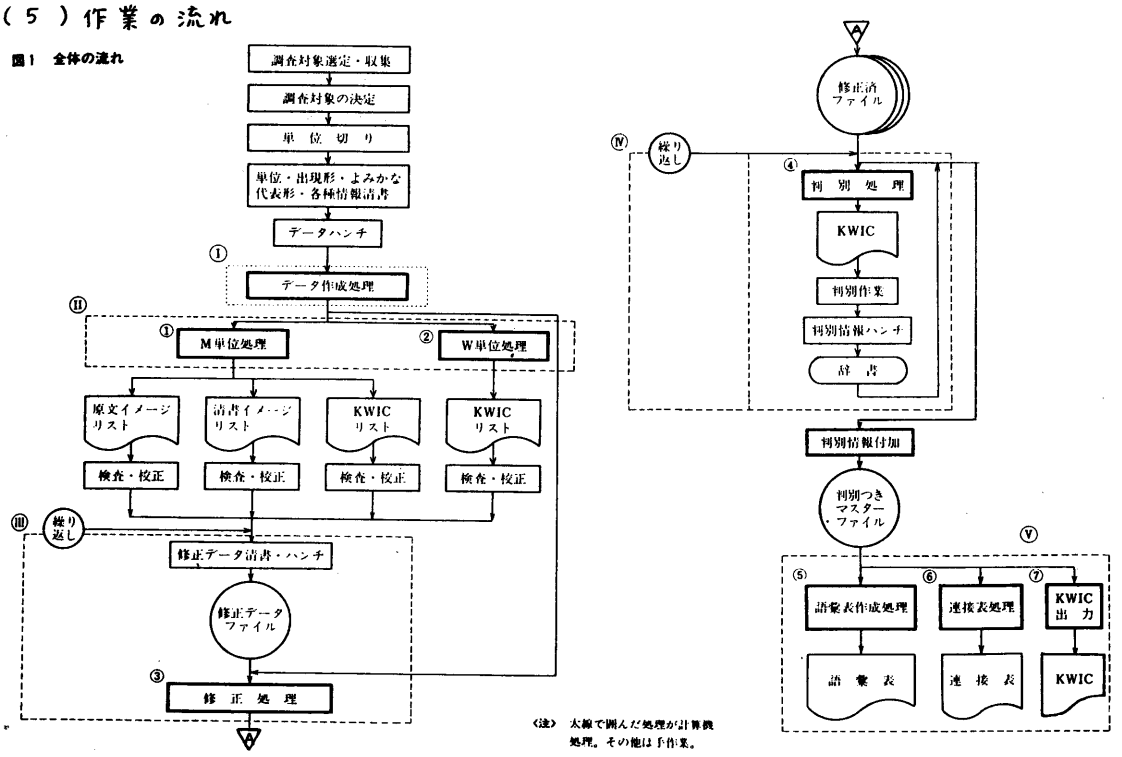
11　ひらく(開)　おこる(起・興)　おこなう(行)
　　あらわす(表・著・現)　たてる(立・建)　なる(成)
　　もつ(持)
12　おくる(送・贈)　おく(置)　あう(合・会)
　　かく(書)　はなす(離・放)　かえる(帰・返・回)
13　きる(切)　かわる(変・代)
14　ひく(引)　ゆく(行)
16　たつ(立)　はかる(図・測・計)
18　つくる(作・造)
21　とる(採・取・執・撮)

これらはすべて動詞であり、異なる表記を持つ、多義語が多い。

すなわち、異形態となる要因には、①語形変化と②表記がある。語形変化には用言の活用のほか、音韻変化〔雨(あめ・あま)，六(ろく・ろっ・りく・りっ)，西国(さいこく・さいごく)〕があり、表記には、多義による漢字使用のほか、まぜがき〔骨ぶい，子ども〕、送りがな〔当：当て・あて〕、習慣〔か条・カ条・箇条；アヒル・あひる〕や外国語音の表記の方法〔ケネディ・ケネディー・ケネジー，江西(こうせい・ちゃんしー)・チャンシー〕，歴史的かなづかい〔あはれ・あわれ〕がある。

(5) 作業の流れ

図1　全体の流れ



〈注〉　太線で囲んだ処理が計算機処理。その他は手作業。

作業の流れは、図1に示すとおりである。一見してわかる通り、人手の作業が非常にすくない。機械処理は、①データ作成処理、②M・W単位処理、③修正処理、④判別処理、⑤出力処理の五つに分かれる。いずれも、高速漢字プリンターによる結果の出力と人手による検査と修正を伴なう。

修正は最初は教科毎におこない、後にマージをして全教科をおこなった。その回数は7回であり、修正量は全体の6.1%である。エラーは単位切り，よみ・代表形つけ，出現形（パンクミス），助辞情報つけの順に多い。

機械によって修正されたエラーレコード数（全体）

	第1次	第2次	第3次	第4次	第5次	第6次	第7次	計
入れ換え	22698	2249	351	1253	131	1108	4	27794
挿　入	4933	340	48	419	22	142	0	5904
削　除	3063	186	96	441	47	198	0	4031
計	30694	2775	495	2113	200	1448	4	37729
%	81.4	7.4	1.3	5.6	0.5	3.8	0.0	100

先にも述べた通り、同語異語の判別は、「代表形」によって異形をまとめ、「判別情報」によって同音語を区別した。「判別情報」は最初、表記の頭一文字をあてた。これによって判別できない語は別の文字をあてた。自動的につけられた判別情報を修正した量は全体の3%（「し」「する」は特別処理）であった。下にその例を示す。

図2 高校教科書 M単位 KWIC

代表形	判別情報	教科	ページ	段落		
	帰	倫	095	03	そ，神の愛を信じ，すべてを神に	帰する　ほんとうの人間のありかたを教え
	帰	政	146	03	では，純然たる個人的原因にだけ	帰せ　られる貧困は，比較的少ないのであ
きする	混	期	政074	04	になるから，その発行には慎重を	期さ　なければならない．また，公債を中
	期	政	032	01	れは，1890年（明治23）を	期し　て開かれる帝国議会に対応するため
	期	日	249	03	年後の1890年（明治23）を	期し　て国会をひらくことを公約した．こ
	期	日	320	00	た．また土地の売買譲渡の公正を	期する　ために，現地の農地委員会にその
	期	政	029	02	院は，もっぱら議案審議の慎重を	期する　という趣旨から設けられていると
きせい	正	政	047	00	けるため，公職選挙法や政治資金	規正　法で，運動期間や運動費用などに制
きせい	気	学	059	02	性物質（ガス）から生じるものを	気成　鉱床（Sn，Mo，Wが多い）（p
きせい	期	日	269	02	97年（明治30）には労働組合	期成　会が生まれ，またこのころから賃金
	期	日	297	01	成立し，さらにこれは婦人参政権	期成　同盟へと発展した．また1917年
きせい	期	日	249	02	，1880年（明治13）に国会	期成　同盟と名称を改め，運動を国会の即
	機	倫	060	01	もできる．この心のしくみは防衛	機制　とよばれるもので，欲求不満を心の
	機	倫	060	01	自己調整するしくみである．防衛	機制　とは，フロイトによって用いられる
	機	倫	060	02	不満の解消をはかろうとする防衛	機制　のひとつで，合理化といわれる．ド
	機	倫	060	03	きるのは，同一視とよばれる防衛	機制　のはたらきがあるからである．自分
	機	倫	061	00	，昇華とよばれる．これらは防衛	機制　の2，3の例である．われわれは，
	機	倫	061	02	中心にあるのが自我である．防衛	機制　は，欲求不満に直面したとき，この
	機	倫	061	00	求の満足を代理させる補償という	機制　もある．失恋した詩人がすぐれた作
きせい	規	日	012	00	集落のなかではある程度の集団的	規制　があったと思われる．しかし，個々
	規	政	131	01	される労働力の経済的条件などが	規制　されるのである．現在日本では，雇
	規	倫	172	02	度や文化などの意識のありかたが	規制　される．ところで生産力はたえず発
	規	政	169	02	いが，それでも諸国家間の関係を	規制　し，国際社会の秩序を保つうえで非
	規	倫	085	01	慣習として伝えられ，社会秩序を	規制　するものにすぎなかった礼の精神的
	規	倫	014	00	解決させるというような意味での	規制　と，産業化優先の政治から，国民の

2．調査結果

（1）語彙表　　語彙表には、頻度・使用率・順位のほかに、見出しの表記等によって語種を、分類語彙表の番号によって品詞と意味番号を、表記例・注記によって表記や同語の範囲などを示した。全体表と教科別、度数順と50音順がある。

意味番号	見出し	表記例・注記	全体 度数	比率	順位	理科 物理	化学	生物	地学	計 度数	比率	社会 倫社	政経	日史	世史	地理	計 度数	比率
1202	ひと	人	629	1.959	46.0	22	2	29	3	56	.51′	292	112	78	53	38	573	2.71
11770	なか	中	620	1.931	47.0	44	38	61	53	196	1.78	126	56	90	121	31	424	2.01′
23060	かんがえる	考える	616	1.919′	48.0	135	38	58	71	302	2.75	192	54	34	11	23	314	1.49′
11742	チュウシン	中心	614	1.912	49.0	35	3	18	38	94	.86′	41	38	127	117	197	520	2.46
23091	しめす	示す	611	1.903	50.0	195	110	62	44	411	3.74	50	44	39	28	39	200	.95′
13115	ズ	図	610	1.900′	51.0	250	53	42	200	545	4.96		10	2	1	52	65	.31′
1360	セイジ	政治	608	1.894′	52.0							34	253	176	102	43	608	2.88′
2367	せる		603	1.878	53.0	66	42	50	19	177	1.61	117	58	128	106	17	426	2.02′
11720	チイキ	地域	601	1.872′	54.0	2	1		73	76	.69	45	39	22	48	371	525	2.49′
1330	ブンカ	文化	597	1.859	55.0			1		1	.01′	57	29	259	119	132	596	2.82

（２）語彙量　　各教科の語彙量と累積使用率分布曲線を下に示す。

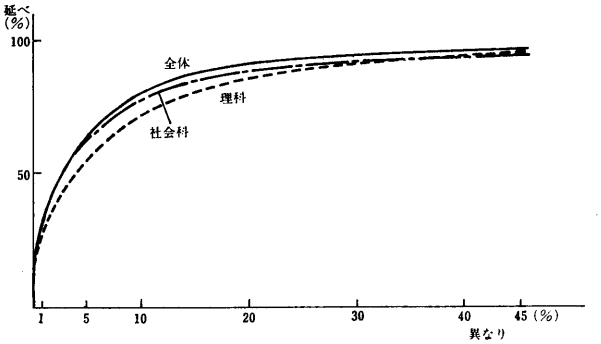
延べ　(%)

	自立語	助辞	数字	記号	全体
物理	29724 (52.25)	18111 (31.84)	1908 (3.35)	7144 (12.56)	56887 (100.00)
化学	26471 (52.07)	14626 (28.77)	2728 (5.37)	7009 (13.79)	50834 (100.00)
生物	30331 (55.67)	17878 (32.81)	858 (1.57)	5418 (9.94)	54485 (100.00)
地学	23288 (53.69)	12725 (29.34)	1649 (3.80)	5714 (13.17)	43376 (100.00)
理科	109814 (53.42)	63340 (30.81)	7143 (3.47)	25285 (12.30)	205582 (100.00)
倫理社会	37875 (54.63)	23972 (34.57)	392 (0.57)	7095 (10.23)	69334 (100.00)
政治経済	44264 (54.58)	25809 (31.82)	2251 (2.78)	8774 (10.82)	81098 (100.00)
日本史	49977 (53.44)	30860 (33.00)	3007 (3.22)	9679 (10.35)	93523 (100.00)
世界史	45130 (54.13)	24892 (29.86)	3462 (4.15)	9883 (11.85)	83367 (100.00)
地理	33998 (55.41)	17716 (28.87)	1445 (2.35)	8203 (13.37)	61362 (100.00)
社会科	211244 (54.35)	123249 (31.71)	10557 (2.72)	43634 (11.23)	388684 (100.00)
全体	321058 (54.03)	186589 (31.40)	17700 (2.98)	68919 (11.60)	594266 (100.00)

異なり　(%)

	自立語	助辞	数字	記号	全体
物理	1804 (95.20)	48 (2.53)	22 (1.16)	21 (1.11)	1895 (100.00)
化学	1742 (95.66)	43 (2.36)	18 (0.99)	18 (0.99)	1821 (100.00)
生物	2573 (97.20)	44 (1.66)	13 (0.49)	17 (0.64)	2647 (100.00)
地学	2777 (97.30)	44 (1.54)	21 (0.74)	12 (0.42)	2854 (100.00)
理科	5205 (98.13)	51 (0.96)	26 (0.49)	22 (0.41)	5304 (100.00)
倫理社会	3770 (97.22)	83 (2.14)	10 (0.26)	15 (0.39)	3878 (100.00)
政治経済	3790 (97.66)	54 (1.39)	23 (0.59)	14 (0.36)	3881 (100.00)
日本史	7091 (98.91)	54 (0.75)	10 (0.14)	14 (0.20)	7169 (100.00)
世界史	5057 (98.60)	42 (0.82)	14 (0.27)	16 (0.31)	5129 (100.00)
地理	4035 (98.29)	44 (1.07)	10 (0.24)	16 (0.39)	4105 (100.00)
社会科	13041 (98.98)	88 (0.67)	23 (0.17)	23 (0.17)	13175 (100.00)
全体	15519 (99.09)	88 (0.56)	28 (0.18)	27 (0.17)	15662 (100.00)

図3. 累積使用率分布曲線(自立語)



表のとおり、全体で約60万語、理科と社会科の割合がほぼ1：2である。異なり語数はほぼ1：3になる。

教科別でみると、最も延べ語数の小さい教科は「地学」、大きい教科は「日本史」で2倍以上である。異なりが最も小さい教科は「化学」であり、「地学」は理科の中で最も大きい。「日本史」は全教科の中でも最も異なり語数が大きく、「化学」の約4倍である。この点から考えると「日本史」は語で表わされる内容が重要であり、「化学」等は語と語の関係で表わされるものが重要であると推測される。

右の図は各教科の意味別語彙量を示したものである。

意味分類は次のとおりである。

.1　抽象的関係

.2　人間活動の主体

.3　人間活動 — 精神および行為

.4　人間活動の生産物 — 結果および用具

.5　自然 — 自然物および自然現象

「化学」をはじめ理科では抽象的関

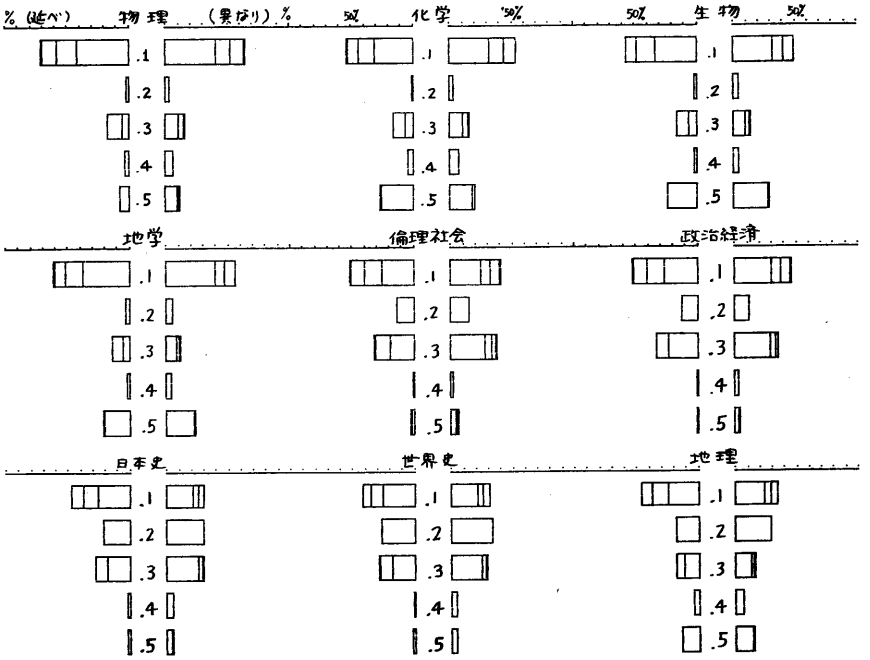


図4. 意味・品詞別語彙量

係（特に「物理」で顕著）・自然が多く、「日本史」をはじめ社会科では、人間活動の主体・人間活動（特に「倫理社会」「政治経済」で顕著）が多い。先に推測したことを裏づけている。さて、この中で「地理」は自然が多い点で理科と似ている。そこで、各教科の共出現語の使用率によって、語彙の類似度を計算してみよう。計算式は、水谷静夫の考案した類似度Dによる。

	物理	化学	生物	地学	倫社	政経	日本史	世界史	地理	異なり語数
物理		.815	.816	.855	.703	.731	.712	.697	.742	1804
化学	.812		.828	.828	.572	.602	.598	.570	.633	1742
生物	.687	.682		.721	.637	.633	.631	.604	.641	2573
地学	.665	.635	.661		.612	.617	.675	.626	.775	2777
倫社	.619	.607	.662	.666		.865	.875	.861	.759	3770
政経	.542	.551	.579	.621	.844		.909	.883	.796	3790
日本史	.448	.450	.485	.526	.710	.735		.795	.669	7091
世界史	.409	.409	.443	.500	.750	.790	.875		.758	5057
地理	.447	.448	.478	.659	.638	.693	.777	.780		4035

教科書間の語彙の類似度　D(s,t)：共通する語の教科sにおける使用率の和

この表をみると理科の間、社会科の間でのD値は当然のことだが大きい。しかし、その中でも「地学」は「地理」との類似度が他の教科とのそれよりも大きい。

（3）各教科高頻度語彙　各教科別度数順語彙表において100位以内の語を集めると次のようになる。

467語　和語117，漢語305，外来語10，地名17，混種語2，英字16
体376，用45，相17，その他3，〃

これらは理科・社会科の重要語彙といえる。そのうち、6教科以上で共に100位以内に入った28語は、「的・様・する・ある・れる・いる・なる・事・この・その・よる（由・因）・られる・これ・それ・また／いう・つくる／もの・できる・ため・ない・みる・とる／日本・化・第・時・持つ」であって、これらは専門的知識をあらわすためだけではなく、日本語として重要な語彙といえる。

しかし、467語のうち、全教科に1回以上用いられた語138語の中には、それ自身は専門的な事柄・現象・概念を表わすものではないが、そのような事柄等を表現するにはなくてはならぬ語が多くみられる。その一部を示す。

的・様・年・化・第・者・性・量・体・力・中・間・数・物・法・式・後・上・点・内・線・層・運動・中心・変化・世紀・関係・発展・必要・状態・問題・独立・化学・方向・自然・一定・次第・位置・発生・利用・実験・核・現象・部分・意味・内部・時間・単位・現在・構造・結合・よる（由・因）・ため・しかし・・更・就く・於ける・共・はじめ・もと・そして・作る・持つ・行なう・起(興)る・受ける・従う・出す・付く・結ぶ・得る・求める・入れる・含む・立つ・あらわす・加える・はかる・はじめる・対する・生ずる・多い・大きい・強い・良い・

逆に、次の語は1教科だけに使われ、かつ100位以内にはいった語である。これらは専門的知識そのものを示す重要語といえよう。

〔物理〕ベクトル・v・ばね　〔生物〕ホルモン・血液・胚・生殖・分泌・腺・変異　〔地学〕マグマ・恒星・星雲　〔政経〕保険　〔日本史〕京都・大名　〔世界史〕朝

国立国語研究所報告76「高校教科書の語彙調査」（秀英出版，昭58年3月）による。