

国立国語研究所学術情報リポジトリ

国語辞典編集のための用例データベース

メタデータ	言語: jpn 出版者: 公開日: 2021-06-11 キーワード (Ja): キーワード (En): 作成者: 木村, 睦子 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003299

国語辞典編集のための用例データベース

国語辞典編集室 木村 睦子

要旨：

既存の辞書に依存しないオリジナルな辞書を作るためには、まず生の用例をたくさん集める必要がある。国語辞典編集室では、昭和54年に準備室が発足して以来、用例採集のための目録作り、作業手順の検討などを行うとともに、国定読本を資料として試験的に用例採集作業を開始した。昭和64年に正式の室になってからは、雑誌を対象に本格的な用例採集を開始し、続いて文学作品・国定算数教科書などに手をつけた。以下にこれまでの作業の概要を述べるとともに、最新の成果である国定算数教科書KWICについて、二三の具体的な事象を述べる。

なお、当編集室では、研究室公開も行うので、その際に閲覧を希望する資料があれば、応じる用意がある。

キーワード：用例辞典 語誌 コンコーダンス KWIC 全数式 スカウト式

1. 日本語用例辞典の構想

(1) 歴史的国語大辞典であること

個々の見出し語について、意味用法・語形等の変遷がたどれるような辞書を作りたい。

(2) 用例中心の辞書であること

歴史的変化をたどるとなると、内省というものは役に立たない。したがって、現代から上代に至るまでの用例を多数採集し、整理した上で収録する必要がある。意味記述は用例からの帰納を原則とする。

2. 用例データベース作成の方法

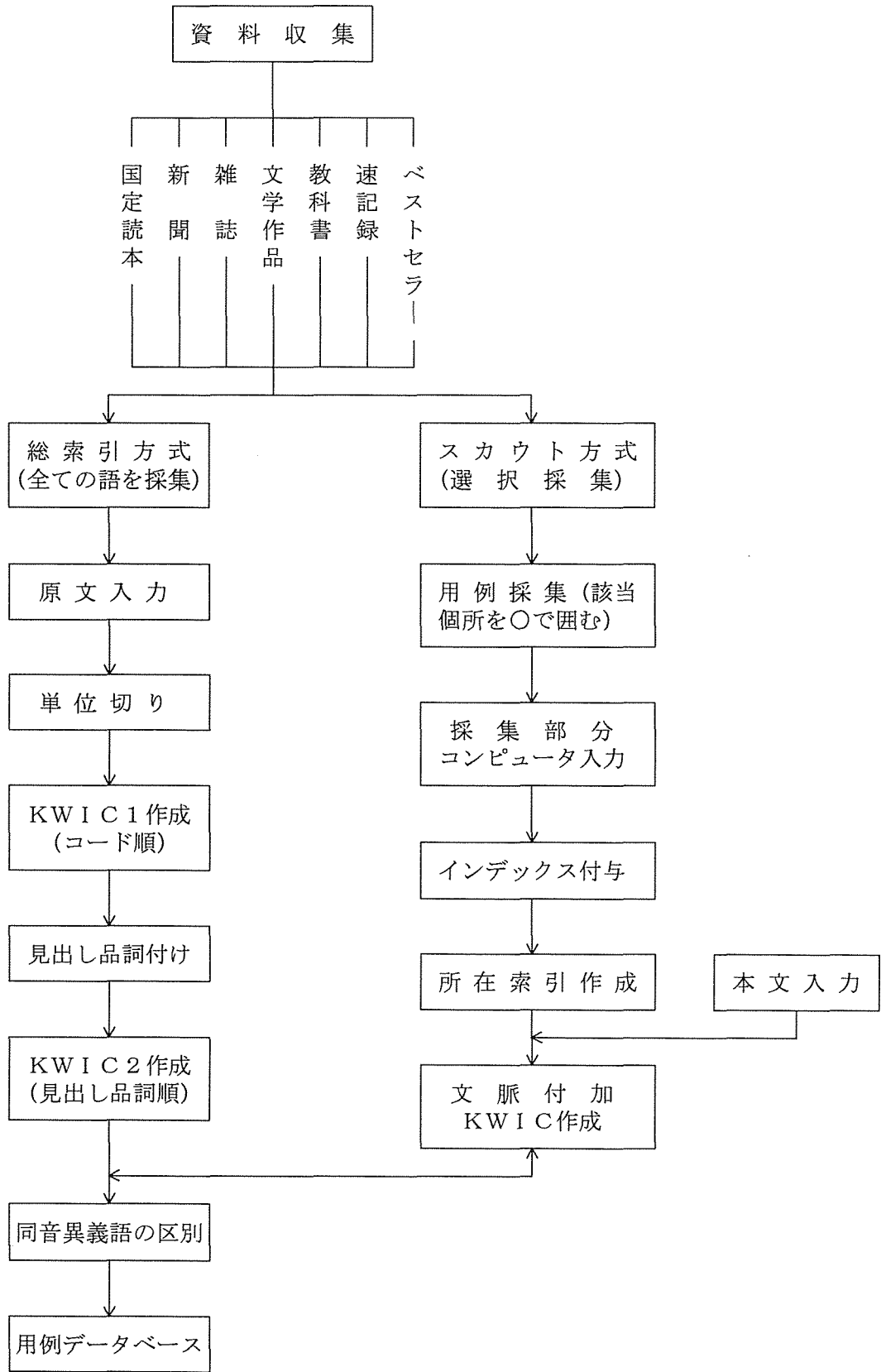
2. 1 全体構想

図1「用例データベース作成の流れ」に示すように、用例採集の方法として、全数式（総索引方式）とスカウト方式（選択採集方式）の二つを考え、対象資料の性質によって採集方法を使い分けるという方針をとってきた。すなわち、全数式においては、すべての用例を網羅して、見出しの頻度はもちろん、意味用法区分ごとに頻度をはかることができるようにする。スカウト方式とは、人間が大量の資料に目を通し、採録したい語を抽出する方式をいい、異なり語数をふやすことと、めずらしい用法を拾うことを目的とする。

我々が手本として見ている外国語の大辞典にOED (Oxford English Dictionary) とフランス語宝典 (Trésor de la Langue Française 略称TLF) とがある。OEDの用例採集法はすべてスカウト式であり、多数のボランティアの助けを借りた人海戦術によるものであった。採集した用例数は600万、辞書に記載した用例数が241万といわれる。他方TLFは、コンピュータによる全数調査方式をとり、1億の用例を持つという（辞書は未完）。選択採集による600万と全数式による1億とどちらが多いかの判定はむずかしい。

我々の方法はいわば折衷方式である。すなわち、調査対象となる資料によって、全数式とスカウト方式とを使い分ける。TLFより遅れてスタートしたにもかかわらず、漢字入力の問題があるために、フランス語の場合ほどコンピュータの威力が発揮できないと考えたのが、その理由である。とはいえ、わずか10年前と比べても、情報機器の発達はめざましく、この面で見直しをせまられている。最初の資料となった国定読本の文脈付き総索引作りは、昭和55年に手作業で開始し、途中から汎用計算機に切り換え、それからさらにパソコンに切り換えて、手作

図 1 . 用例データベース作成の流れ



業でやった部分を作り直すという煩雑な手順をふむこととなった。他の資料についても、そういう面がないとはいえない状況にある。

2. 2 調査対象資料

用例採集の対象は、現代から始めて上代にさかのぼるという方針で、当面の目標を1901（明治34）年から1950（昭和25）年までとした。この時期をえらんだ理由は、その時期に日本語の標準語が成立したと思われるからである。その期間における雑誌、文学作品などの目録として、以下の三つができています。これらの目録は、単に雑誌名や作品名を挙げただけでなく、それらに語彙資料としての評点を与えている。

用例採集のための主要文学作品目録 昭和58 国語辞典編集準備資料 2

用例採集のための主要雑誌目録 昭和58 国語辞典編集準備資料 3

用例採集のためのベストセラー目録 昭和59 国語辞典編集準備資料 4

ただし、国語辞典編集室の規模からして多くの作業をこなすことは不可能であり、実際に手を付けたものは、着手した順に以下の4種である。

- ① 国定読本（全数式）
- ② 総合雑誌『太陽』（スカウト式）
- ③ 文学作品（全数式）
- ④ 国定算数教科書（全数式）

2. 3 作業の進行状況

2. 3 1 国定読本

（1）資料の性格

国定読本というのは、明治37年4月から昭和24年3月までの間に使用された文部省著作の小学校用国語教科書6種のことである。その6種を使用時期に従って示すと下記の通りである。

第1期 明治37年より使用『尋常小学読本』（通称イエスシ読本）8巻

第2期 明治43年より使用『尋常小学読本』（通称ハタタコ読本）12巻

第3期 大正7年より使用『尋常小学国語読本』（通称ハナハト読本）12巻

第4期 昭和8年より使用『小学国語読本』（通称サクラ読本）12巻

第5期 昭和16年より使用『ヨミカタ』『よみかた』『初等科国語』（通称アサヒ読本）12巻

第6期 昭和22年より使用『こくご』『国語』（通称みんないいこ読本）15巻

底本はおおむね初年度使用本である。

（2）作業経過

上記の本文すべてを単位切りし、各単位ごとに見出し・品詞・層別情報・文脈等をつけ、見出しの五十音順、品詞番号順などによって配列したのが『国定読本用語総覧』1～11（三省堂）である。文脈の範囲は人手によってきめたが、5期から一部（助詞・助動詞など）をKWIC化した。これらの作業は1期から順におこなわれ、逐次書物として刊行された。そこで全体を見渡せるように、1～6期のすべての見出しに、期ごとの頻度と合計を付与したのが『国定読本用語総覧』12（総集編）である。さらにすこし遅れて、これらの用例をすべてKWIC形式にして、CD-ROMで刊行した。

見出し総数3万2千、延べ用例数は60万である。

2. 3 2 雑誌『太陽』

(1) 資料の性格

『太陽』は明治28年から昭和3年まで33年間にわたって博文館から刊行された月刊の総合雑誌である。出版マスの先駆といわれ、隆盛を誇っていた博文館が、当時刊行中だった5種の雑誌（それぞれ評論・商業・農業・法律・婦女の専門誌）をことごとく廃刊し、『太陽』一本にまとめたものである。

1901（明治34）年から始めて8年ごとに1年分（臨時増刊号を除く）をとり、それに創刊年と廃刊年（2号のみ）の分を加えて、計62冊を調査対象とした。

国語研究所に原本がそろっているのを、それを底本にした。

(2) 作業経過

読本とちがって選択採集方式なので、まず人間が本文に目を通し、拾うべき語句に印をつけ、それを出典情報とともに入力する。採集した文字列そのままでは検索が困難なので、それにインデックスと称する読みを付与してインデックス・ファイルを作成する。一つの用例に複数のインデックスがつくこともある。さらに近年に至って、インデックスに文脈を付加してKWIC化するための作業が始まった。

インデックス付与作業は一わたり終了したが、読みを付け終えただけで、同音異語の判別などは、まだ1年分しかできていない。文脈付加のためのコーパス作りは、約1年分がおわったところである。3、4年後の電子媒体による刊行をめざしている。

インデックスは全体で延べ70万語である。そのうち、1901年の分は、延べ語数15万5千に対して、異なり語数が6万1千である。延べ語数にくらべて異なり語数が大きいのは、選択採集方式によるところが大きい。

2. 3 3 文学作品

(1) 資料の性格

対象とする文学作品は国語辞典編集準備資料2「用例採集のための主要文学作品目録」に挙げられている1901年～1950年に発表された文学作品のうち、所外の10名の専門家の方々によって選ばれた主要文学作品 116点を当初の対象とした。その後、

- ・ より幅広く語の異なり用例を集める
- ・ 作家／作品による特定の語の出現数偏向を避ける

という観点から、116点の作品のうち、概算で1万文節以下のものを当面の対象とし、さらに「用例採集のための主要文学作品目録」作成の準備段階資料である「用例採集のための文学作品選定目録」から、やはり概算で1万文節以下の作品を優先して調査対象としていくこととした。

(2) 作業経過

作業方法としては国定読本に対する作業とほぼ同様であり、入力された本文を単位切りしてKWIC化し、そこに、検索語、品詞、層別情報といった情報を付与していく。読本と異なるのは、形態素レベルに分割した検索語がつくことである。

現在、50余作品が本文入力され、上記作業が進行している。またそれと並行して雑誌『太陽』との統一的な単位の規準、旧漢字の処理についての規準などを設けるための検討をしている。

2. 3 4 国定算数教科書

(1) 資料の性格

国定算数教科書も読本と同じく1期から6期までであるが、1期は教師用のみで児童用がないので、2期以降を調査対象とした。底本として『日本教科書大系』（講談社）を用いた。この

教科書大系は教科によって扱いが異なり、部分的に活字化しているところがあるが、算数はすべて写真版である。

算数であるから、当然練習問題がたくさんあるが、そういうものを除き、文章の体をなしている部分を調査対象とした。したがって読本より量が少なく、3分の1程度である。

第2期	『尋常小学算術書』	3～6学年	4巻	明治43.	3
第3期	『尋常小学算術書』	3～6学年	4巻	大正 8.	3
第4期	『尋常小学算術書』	1～6学年	12巻	昭和 9.	12
第5期	『カズノホン』『初等科算数』	1～6学年	12巻		16. 3
第6期	『さんすう』『算数』	1～6学年	9巻		22. 3

(2) 作業経過

作業手順は読本とほぼ同じであるが、単位の切り方や見出しの立て方、数詞の扱い方など、いくらか読本と異なるところがある。たとえば、読本では文語と口語で代表語形や活用型の異なる用言は別見出しになっているが、ここでは口語形に統一した。また数字（漢数字・アラビヤ数字とも）の列は、ほとんどすべて○に置き換えた。2～6期全体のKWICと語彙表が一応完成している。

延べ語数20万、異なり語数6600である。

3. 国定算数教科書における語彙の変遷

算数の語彙には、四則演算・図形・度量衡など、教科として教える必要のある言葉のほかに、応用の場として用いられるものが多く出現する。数量は日常生活のさまざまな部分に関係するが、木村は、そのうちのどれがテーマとして多く用いられているか、またどのように表わされているか、期によってどのように違うかに興味をもった。まだ全体を分類し、見渡すに至らないが、二三の特定の領域を選んで考察した。結果として、期による違いが読本よりはるかに大きいことがわかった。

3. 1 四則演算に関する用語

国定算数教科書に出てくる四則演算に関する用語のうち、おもな名詞と動詞を表1に掲げた。以下にその特徴をあげる。

- ① 加算を表わす動詞には「足す」「寄せる」「加える」の三つがあつて、2・3期では「足す」が優勢であるが、4・5期は「寄せる」が多数を占め、6期には「寄せる」「加える」がそれぞれ半数近くを占める。

4期・5期において例外的に使われた「足す」の例は以下の通りである。

(例) $3+1$ ヲ 「3 タス 1」 ト ヨミマス。

6期も2例中1例は同様のものである。+記号を「足す」と読ませておきながら、地の文では「寄せる」または「加える」を使っている。また足算という語は出現しない。

- ② 掛算・割算はそうでもないが、寄算・引算においては、2期と3期は動詞の使用が少なく、その代りに名詞「和」「差」が多く用いられている。
- ③ 「加法・減法・乗法・除法」または「寄算・引算・掛算・割算」という言い方は4期以降減少傾向にあり、特に6期において顕著であるが、その代りに「計算」という上位語が飛躍的に増えている。これは、「次の加法を行へ」「次の割算をなせ」などという言い方が、「次の計算をなさい。」などと変化したせいである。

3. 2 金銭関係の用語

国定算数教科書に出てくる金銭関係の用語のうち、おもな名詞と動詞を表2に掲げた。これらの見出し語については、個々の用例をチェックして、用法の異なるものは除外した。したがって、ここにあげた頻度はすべて金銭に関するものである。

用語を内容によって次の5種に分類した。

①金銭一般	12語 (名詞 7 動詞 5)
②税金関係	22語 (名詞 21 動詞 1)
③給与関係	7語 (名詞 7)
④預貯金・株	16語 (名詞 13 動詞 3)
⑤売買関係	17語 (名詞 13 動詞 4)

全般に期がくだるにしたがって、使用頻度が減っており、2期と6期では比率にして1桁以上の差がある。特に税金関係においてその傾向が著しい。

表1. 四則演算に関する語彙と頻度
() 内は百分率

		2期	3期	4期	5期	6期	計
かほう	加法	23	27	0	0	0	50
げんぼう	減法	26	28	0	0	0	54
じょうほう	乗法	34	31	0	0	0	65
じょほう	除法	31	36	0	0	0	67
よせざん	寄算	44	61	36	35	17	193
ひきざん	引算	39	60	33	33	15	180
かけざん	掛算	61	87	56	52	19	275
わりざん	割算	70	95	66	49	15	295
わ	和	10	13	2	0	2	27
さ	差	16	11	3	7	5	42
せき	積	11	1	3	0	3	18
しょう	商	13	1	4	0	1	19
(小計)		378 (2.08)	451 (2.13)	203 (0.36)	176 (0.36)	77 (0.12)	1285 (0.62)
くわえる	加	1	3	3	4	18	29
たす	足	3	8	1	1	2	15
よせる	寄	1	0	63	34	16	114
ひく	引	2	1	46	58	62	169
かける	掛	9	8	26	39	51	133
わる	割	10	16	33	32	28	119
(小計)		26 (0.14)	36 (0.17)	172 (0.31)	168 (0.35)	177 (0.28)	579 (0.28)
けいさん	計算	33	23	89	79	429	653
けいさんする	計算	48	33	108	77	205	471
(総計)		485 (2.67)	543 (2.56)	572 (1.02)	500 (1.03)	888 (1.43)	2988 (1.45)

表2. 金銭関係の語彙と頻度

() 内は百分率

① 金銭一般		2期	3期	4期	5期	6期	計
〇えん	一円	132	155	86	37	35	445
〇えん〇せん	一円一銭	67	78	69	38	22	274
きん〇えん	金一円	15	18	3	0	0	36
かね	金	8	11	13	3	0	35
おかね	御金	0	0	113	47	18	178
きんがく	金額	4	2	2	0	3	11
きんだか	金高	12	12	3	2	6	35
しはらう	支払	0	0	0	0	8	8
はらう	払	8	12	43	5	2	70
える	得	35	24	16	3	2	80
うけとる	受取	4	4	1	1	0	10
もらう	貰	2	1	5	1	2	11
(小計)		287 (1.58)	317 (1.50)	354 (0.63)	137 (0.28)	98 (0.16)	1193 (0.58)
② 税金関係		2期	3期	4期	5期	6期	計
ぜいきん	税金	7	0	0	0	0	7
そぜい	租税	0	1	1	0	0	2
そぜいこうか	租税公課	0	0	1	0	0	1
ちそ	地租	12	9	0	0	0	21
しょとくぜい	所得税	14	11	0	0	0	25
ほんぜい	本税	0	1	0	0	0	1
ふかぜい	付加税	0	4	0	0	0	4
しょとくぜいふかぜい	所得税付加税	0	1	0	0	0	1
しぜい	市税	0	3	0	0	0	3
ふぜい	府税	0	3	0	0	0	3
えいぎょうぜい	営業税	4	2	0	0	0	6
かおくぜい	家屋税	0	4	0	0	0	4
つうこうぜい	通行税	0	1	1	1	0	3
いんしぜい	印紙税	1	0	0	0	0	1
しゅぞうぜい	酒造税	0	2	0	0	0	2
ゆうぜい	郵税	2	0	0	0	0	2
ぜいりつ	税率	24	6	0	0	0	30
しょとくぜいりつ	所得税率	1	2	0	0	0	3
ちそりつ	地租率	0	2	0	0	0	2
かぜいがく	課税額	0	3	0	0	0	3
かぜいきんがく	課税金額	0	3	0	0	0	3
おさめる	納	20	18	0	0	0	38
(小計)		85 (0.47)	76 (0.36)	3 (0.01)	1 (0.00)	0 (0.00)	165 (0.08)
③ 給与関係		2期	3期	4期	5期	6期	計
ほうきゆう	俸給	1	1	2	2	4	10
きゆうりょう	給料	2	0	0	1	0	3
きゆうきん	給金	0	4	2	1	2	9
げつきゆう	月給	3	2	3	3	0	11
につきゆう	日給	0	3	4	3	4	14
ちんぎん	賃金	2	0	3	0	2	7
ちんせん	賃銭	15	13	8	2	0	38
(小計)		23 (0.13)	23 (0.11)	22 (0.04)	12 (0.02)	12 (0.02)	92 (0.04)

④ 預貯金・株		2期	3期	4期	5期	6期	計
ちよきん	貯金	2	1	9	8	0	20
よきん	預金	0	0	2	0	0	2
がんきん	元金	8	21	15	0	0	44
がんりごうけい	元利合計	8	9	10	0	0	27
りし	利子	10	16	39	0	0	65
りそく	利息	14	28	0	0	0	42
りりつ	利率	10	6	10	0	0	26
きんり	金利	1	0	0	0	0	1
り	利	2	8	0	0	0	10
○かぶ	一株	17	11	0	0	0	28
かぶ	株	6	3	0	0	0	9
かぶしき	株式	0	4	0	0	0	4
はいとう	配当	5	4	0	0	0	9
あずける	預	3	2	29	8	2	44
ちよきんする	貯金	0	0	7	8	0	15
ためる	貯	1	1	0	1	0	3
(小計)		87 (0.48)	114 (0.54)	121 (0.21)	25 (0.05)	2 (0.00)	349 (0.17)
⑤ 売買関係		2期	3期	4期	5期	6期	計
だいか	代価	5	0	13	11	0	29
だいきん	代金	12	3	0	1	0	16
ねだん	値段	2	0	40	9	5	56
かかく	価格	0	0	7	1	0	8
かがく	価額	0	0	12	0	0	12
おつり	御釣	0	1	17	8	5	31
つりせん	釣銭	4	0	0	0	0	4
りえき	利益	5	9	0	0	0	14
りえききん	利益金	7	0	0	0	0	7
そん	損	2	2	0	0	0	4
そんえき	損益	4	1	0	0	0	5
そんきん	損金	1	0	0	0	0	1
そんしつ	損失	0	3	0	0	0	3
うる	売	31	18	17	11	0	77
かう	買	40	35	146	55	10	286
もうける	儲	6	5	10	0	7	28
そんする	損	3	3	0	0	0	6
(小計)		122 (0.67)	80 (0.38)	262 (0.46)	96 (0.20)	27 (0.04)	587 (0.28)
(総計)		604 (3.32)	610 (2.87)	762 (1.35)	271 (0.56)	139 (0.22)	2386 (1.16)
(総延べ語数)		18174	21193	56353	48398	62259	206377