国立国語研究所学術情報リポジトリ

国立国語研究所新聞記事データベース：
ことばに関する記事1949年～現在

メタデータ	言語: Japanese 出版者: 公開日: 2021-06-11 キーワード (Ja): キーワード (En): 作成者: 池田, 理恵子, 辻野, 都喜江 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003328

国立国語研究所新聞記事データベース
—ことばに関する記事　1949年～現在—

情報資料研究部第一研究室
池田　理恵子・辻野　都喜江

要旨:
国立国語研究所では，創立直後の1949年から現在にいたるまで，ことばに関連する内容の新聞記事を収集し，『新聞所載国語関係記事切抜集』（『切抜集』）として保存している（総数約11万件）。『切抜集』は戦後の日本語及び日本人の言語生活の変遷の一端を物語る貴重な資料であるが，収録記事数が多い上，体系的な記事目録がないため，そのままの形では研究資料として活用することは困難であった。国立国語研究所情報資料研究部では，この蓄積記事について，掲載日，掲載紙名，見出し等を収録し，記事検索に有用な情報を付加した『国立国語研究所新聞記事データベース』（『データベース』）を作成してきた。現在までに，蓄積記事のすべてについて，記事検索情報の付与がほぼ完了した。
本報告では，『データベース』について，データの概要を示し，次いで，過去約50年を対象として，年別件数と分野別件数の推移，記事の執筆者の属性別件数について概観する。戦後の日本人のことばに対する関心のあり方や，日本語あるいはことばに対する価値観を探る社会言語学的な史的研究に対して，研究の視点や手がかりを提示する。

キーワード:　新聞記事　データベース　日本語　国語

1.新聞記事を利用した研究と『国立国語研究所新聞記事データベース』

国立国語研究所では，創立直後の1949年から，ことばに関連する内容の新聞記事を収集し，『新聞所載国語関係記事切抜集』（以下，『切抜集』）として保存している。その一部は『国語年鑑』の「新聞記事一覧（国語関係）」に掲載されている（1989年版まで）。『国語年鑑』の「展望の部」（1988年版まで）では分野別に一年間の動きが示されているが，その中には新聞記事を資料として用いながら動向を述べているものもある。また，新聞記事を資料として，特定のテーマについてことばに対する意識（の変遷）を考察したものに尾崎(1995)や田上(1992)がある。

近年，新聞記事データベースは提供と利用の両面で飛躍的な伸びを遂げている。また，インターネットを利用して日々のデータを得ることも可能になっている。しかし，新聞界で記事データベース構築の機運が高まったのは1980年代の半ばからであり，日々新たな記事が追加され，記事データベースが更新されるなかにあって，（データベース化に着手した時点より）過去に遡って記事を収録するのは至難の技である[1]。その上，それらの記事データベースは，基本的に汎用データベースであるために，「日本人の言語及び言語生活の研究」という特定の目的のために利用することは難しい。その点，『切抜集』は「国語・日本語」「言語生活」というテーマのもとに収集された資料であり，かつ，50年という極めて長い期間をカバーするものである。

しかし，『切抜集』は収録記事数が多い上，体系的な記事目録がないため，そのままの形では研究資料として活用することは困難であった。そこで，国立国語研究所情報資料研究部では，この蓄積記事資料の有効な活用法の一つとしてデータベース化があると考え，「国語関係新聞記事の蓄積と活用の研究」を行ってきた。現在までに，蓄積記事のすべてについて，記事検索情報の付与がほぼ完了した。

新聞記事を資料とした研究においては，新聞というメディアの特性や記事選択に際しての「新聞社の配慮」を無視することはできないが，ことばをめぐって何が変わり，何が変わらないのか，変わったとしたらどのように変わったのか，『国立国語研究所新聞記事データベース』及び『切抜集』は一つの手がかりを与えてくれるように思う。

2.『国立国語研究所新聞記事データベース』の情報

『国立国語研究所新聞記事データベース』（以下，『データベース』）には以下の情報が入力されている。

①台帳番号（『切抜集』の冊子の通し番号。）
②記事番号（『切抜集』の台紙の通し番号に枝番「0」を加えたもの。）
③年
④月日
⑤新聞名
⑥朝夕（朝刊夕刊の別。）
⑦地方版（全国紙の東京本社版以外の場合。）
⑧ページ
⑨掲載面（「社会」「家庭」等の面の名称。）
⑩執筆者属性（内部：新聞社内部の人が執筆した記事。投書：投書欄の記事。依頼：（新聞記者以外の）専門家が書いた記事。投書・内部：読者の質問に新聞記者が答えている記事。投書・依頼：読者の質問に（新聞記者以外の）専門家が答えている記事。）
⑪氏名（依頼記事の執筆者の氏名。）
⑫欄名
⑬見出し
⑭備考（見出しだけでは記事の内容が把握できない場合の補足的情報。）
⑮キーワード候補（その記事が扱っている分野・内容。）（3.参照）
⑯フリータ－ム（記事の内容を端的に表す語句。記事に使われている語句をそのまま用いることもある。）

3.『データベース』に収録された記事の内容

『データベース』では，その記事が扱っている分野・内容を示す情報をキーワード候補と呼んでいる。この情報は，所定の83項目の中から該当する項目を選んで入力している。その際，記事の内容が複数の分野にわたると考えられる場合は，複数の分野を選んで入力している。紙面の制約等により，本報告ではキーワード候補の各項目の説明は省略し，『データベース』に収録されている記事の内容を，キーワード候補の例を〈〉付きで示しながらおおまかに示す。

Ａ言語・日本語・日本文化についての学術的研究に関するもの　〈日本語研究〉
Ｂ音声・音韻に関するもの（発音，アクセント，イントネーション等）〈音声・音韻〉
Ｃ文字・表記に関するもの（漢字の字体〈漢字〉，〈仮名遣い〉，〈送り仮名〉等）
Ｄ語彙・用語に関するもの（気になる語，〈外来語〉，〈流行語〉等）　〈語彙・用語〉
Ｅ文法に関するもの　〈文法〉
Ｆ文章・文体に関するもの（文章の書き方，比喩表現，作家の文体等）〈文章〉〈文体〉
Ｇ方言に関するもの　〈方言〉
Ｈことばと社会に関するもの（〈新聞〉，〈放送〉，〈敬語〉，〈電話〉，パソコン・ワープロの利用〈情報化社会〉等）
Ｉことば一般に関するもの（日本語の特質，美しい日本語等）　〈日本語〉
Ｊ国語国字問題に関するもの（国語審議会，常用漢字等）　〈言語政策〉
Ｋことばと教育・学習に関するもの（〈国語教育〉，〈日本語教育〉，外国語教育〈外国語習得〉等）
Ｌ国語学的資料に関するもの（木簡・古文書の発見等）　〈言語資料〉
Ｍ言語・日本語関係諸学会の設立等に関するもの　〈学界動向〉
Ｎ国立国語研究所に関するもの　〈国立国語研究所〉

○海外の言語事情に関するもの（日本以外の諸外国におけることばについての記事，諸外国における日本語・日本語教育等に関するものも含む）　〈海外言語事情〉

基本として，キーワード候補は分野を示すものとして設定した。しかし，『データベース』に収録されているデータは，記事の内容によって件数の多いものと少ないものがあり，記事検索をより効率的にかつ容易に行うことができるように，ある分野についてはさらに細かく分けて，現在83項目を設定している。それらの項目は上位・下位といった分け方になっていないため混乱を招く恐れがあり，キーワード候補の各項目の名称については，今後の検討・整備が必要である。

4.収集対象とした新聞の種類と数

収集対象とした新聞は，1989年５月からは「朝日・毎日・読売」の３紙（東京本社版）に限定している。それ以前はブロック紙・地方紙や「図書新聞」「新聞協会報」等の専門紙からも記事の収集をしていた。対象紙数は，収集を開始した1949年は９紙であったが，最も多い年では40紙（1960年）にのぼり，1949年から1989年までの平均紙数は25.3である。その中には国立国語研究所に寄贈された記事が含まれているため，国立国語研究所において定期的かつ継続して収集対象とした紙数はそれよりも少ない。ちなみに，一つの新聞から一年に50件以上の記事を収集した紙数を計算すると，1950年（1949年は総記事数が87件であるため除外）から1989年までの平均紙数は8.8紙である。

5.年による記事数の推移

『データベース』に収録されている記事を年代別に集計した結果が表１である（1998年分はデータ処理中のため除く）。後述のように，1989年５月を境に収集対象紙の数及び記事収集の基準が異なることをふまえて，1949年から1997年の49年を表にあるような５つの年代に分けた（以後，便宜上，1950年代，1960年代，1970年代，1980年代，1990年代と呼ぶ）。記事総数は，110,701件であり，1970年代と1980年代の記事が非常に多い。

表１　年代別の記事数と割合

	1949-1959	1960-1969	1970-1979	1980-1988	1989-1997	計
記事数(件)	11603	17483	27528	37557	16530	110701
	10.48%	15.79%	24.87%	33.93%	14.93%	100.00%

表には示していないが，年単位の記事数の推移をみると，1949年は87件と少ないが，徐々に件数が増加し，1953年には1,000件を超えている。1970年代後半には3,000件を，1980年代には4,000件を超えているが，この間，定期的・継続的に収集対象とした新聞の数はそれほど増えていない。収録記事の増加は，新聞においてことばに関する内容の記事が増えたことを意味するのだろうか。データを見てみると，1980年代には「文字クイズ」「名字百科」等の連載記事が多く収録されている。一つの連載の年間記事数は約50～100件で，この期間の連載記事の多さが記事件数の増加の一要素と考えられる。

1990年代に記事数が大幅に減少しているのは，1989年から購読新聞を３紙に限定したことに加えて，収集する記事の内容をしぼったことによる。1980年代に『データベース』に数多く収録されているような連載記事（クイズや語句の解説のような内容）は，1989年以後も新聞に掲載されているが，それらは辞書類を見ればわかることであろうと判断し収集対象外とした。新聞に掲載されているというそのことが，ことばに対する関心の表れであるという考え方もあろうが，記事の内容としてより直接的にことばに対する意識や関心のあり方が述べられているものにしぼることとした。

6.キーワード候補による記事数の推移

キーワード候補のうち，付与された記事数の多いもの，年代による増減の幅の大きそうなものについて，年代別の記事数とその年代の記事総数に対する割合を表２に示す。

表2 キーワード候補別の記事数と割合

	1949-1959	1960-1969	1970-1979	1980-1988	1989-1997	計
年代別記事数	11603	17483	27528	37557	16530	110701
漢字(計)*1	1460	2134	2890	3309	745	10538
	12.58%	12.21%	10.50%	8.81%	4.51%	9.52%
送り仮名	218	274	278	24	6	800
	1.88%	1.57%	1.01%	0.06%	0.04%	0.72%
仮名遣い	430	336	182	258	29	1235
	3.71%	1.92%	0.66%	0.69%	0.18%	1.12%
語彙・用語	1727	2907	3800	5855	2100	16389
	14.88%	16.63%	13.80%	15.59%	12.70%	14.80%
方言(計)*2	639	1715	1248	1774	579	5955
	5.51%	9.81%	4.53%	4.72%	3.50%	5.38%
言葉遣い	568	901	941	1348	729	4487
	4.90%	5.15%	3.42%	3.59%	4.41%	4.05%
敬語	306	440	658	712	347	2463
	2.64%	2.52%	2.39%	1.90%	2.10%	2.22%
情報化社会	172	477	934	2812	2089	6484
	1.48%	2.73%	3.39%	7.49%	12.64%	5.86%
言語政策	709	1220	1164	777	365	4235
	6.11%	6.98%	4.23%	2.07%	2.21%	3.83%
国語教育	1255	1316	1499	1178	572	5820
	10.82%	7.53%	5.45%	3.14%	3.46%	5.26%
日本語教育	68	254	744	1705	712	3483
	0.59%	1.45%	2.70%	4.54%	4.31%	3.15%
外国語習得	297	698	1504	2236	1245	5980
	2.56%	3.99%	5.46%	5.95%	7.53%	5.40%

*1 漢字(計) キーワード候補「漢字」「漢字制限」のどちらか一つでも付与されたもの。
*2 方言(計) 実際のデータでは地域を分けている。

語彙・用語は付与された件数が非常に多い[2]。割合をみても，どの年代でも10％台で大きな変動はなく，全体でも14.80％で最多である。

語彙・用語の次に付与された件数が多いのは漢字で，全体で9.52％である。年代による変化をみると，1990年代はそれ以前に比べて割合が半減している。1980年代までの件数の多さや全体に占める割合の大きさと1990年代の減少について，『データベース』の見出しの項目をもとに考えると，年代による変化の原因として大きく二つのことが挙げられる。一つは，1980年代までは，漢字についてのクイズや豆知識的な連載記事が多かったことである。もう一つは，1980年代までは，漢字についての国語審議会の報告や答申，内閣告示等に関する記事が多かったこと，及びそれらに対する識者の意見記事や読者の投稿記事が非常に多かったことである。この点については，件数の変化を年単位や答申の前後に着目して考察したり，述べられている意見がどのようなものかを詳細にみていくこともおもしろいテーマである[3]。

送り仮名と仮名遣いは全年代を通じて件数が少なく割合も小さい。その中での細かな変化を見ると，送り仮名については，内閣告示のあった1959年（送りがなのつけ方）と1973年（送り仮名の付け方）を含む時期に比べ，1980年代以後は件数が激減している。仮名遣いについても同様に，内閣告示のあった1946年（現代かなづかい）と1986年（現代仮名遣い）を含む時期を過ぎると，記事数が極めて少なくなっている。

言語政策の付与された記事の割合は1950,1960年代は6％台だが，1970年代は4％台，1980年代以後は2％台と少なくなっている。キーワード候補「言語政策」は，国語審議会の中間報告・答申とそれらに対する意見，さらに，国語審議会や国語政策に望むこと等の内容の記事に付与される。1970年代までは漢字や送り仮名，仮名遣い，ローマ字等について国語審議会で審議されていて，その経過を報告・紹介する記事や，国語審議会の答申や試案に対する解説記事や識者の意見，読者の投書が数多く収録されている。年ごとに件数や割合の変動を見ると，答申や報告の前後に記事が集中している。1980年代以後は，

常用漢字表（1981年）と外来語の表記（1991年）について内閣告示のあった年の前後に件数が増えるものの，それ以外の時期のデータは少ないようである。

方言については，田上(1992)は新聞記事を資料として方言復権の軌跡を考察していて，方言に対する価値観や意識のあり方はこの50年で大きく変化していることが予想される。件数の推移より，記事本文をみていくことで，その変遷を探ることができそうである。

言葉遣いや敬語は，全年代を通してそれぞれ４％前後と２％前後で推移している。一方，情報化社会は年代による変化が大きい[4]。年ごとにデータを見ると，件数は1970年代後半から急激な曲線を描いて増加していて，年代別記事総数に対する割合も1980年代，1990年代ともに，前年代に比べほぼ倍増している。

ことばと教育について，国語教育，日本語教育，外国語習得をみると，国語教育は，年代を通して記事数があまり増加せず，全体の記事数が増加しているため，結果として年代を経るごとに割合が減少している。対照的に，外国語習得は年代を経るごとに記事数も増加し，割合も順に増えている。日本語教育は1950,1960年代は記事数が少なく，1970年代後半から増加している。ただし，外国語習得や日本語教育が付与された記事の件数や割合が年を追うごとに増加傾向にあることを，単純に新聞に掲載される件数が増加したためととらえることはできない。収集対象の範囲が，国語の学習・教育という限定されたものから，日本人のことばの学習・教育，さらに，日本語を母語としない人々の日本語学習・教育にも広げられたことの影響も考慮する必要がある。

7.執筆者の属性によるキーワード候補の記事数の推移

『データベース』の記事は誰が書いたかによって「内部」「投書」「依頼」「投書・内部」「投書・依頼」に分けられる。年ごとにそれぞれの執筆者属性が占める割合をみると，「投書・内部」と「投書・依頼」は全年代を通して割合が極めて小さく，平均ではそれぞれ1,04％と0,24％となる。他の三つの属性の割合は，年によって増減はあるが全体を通してみるとその幅はそれほど大きくないようである。平均では，「内部」55,41％，「投書」16,06％，「依頼」27,23％である。

ここでは，執筆者属性を分けた場合に，読者の意見を最も直接的に反映していると考えられる「投書」に着目して，キーワード候補のいくつかを取り上げ，投稿者がことばのどのようなところに関心を持っているかみてみよう（表３）。

表3 キーワード候補別の投書の件数と割合

キーワード候補	記事数(件)	投書数(件)	投書の割合(%)
記事全体	110701	17780	16.06
漢字(計)	10538	2272	21.56
送り仮名	800	52	6.5
仮名遣い	1235	112	9.07
外来語の表記	606	122	20.13
語彙・用語	16389	2204	13.45
各種用語	3680	350	9.51
外来語	6120	1205	19.69
流行語	3314	282	8.51
方言(計)	5955	668	11.22
言葉遣い	4487	1613	35.95
あいさつ	1902	852	44.79
呼称	2046	867	42.38
敬語	2463	758	30.78
情報化社会	6484	255	3.93
言語政策	4235	743	17.54
国語教育	5820	1280	21.99
日本語教育	3483	225	6.46
外国語習得	5980	1063	17.78

それぞれのキーワード候補が付与された記事にどのくらい投書が含まれているかを見ると，投書の割合が１割に満たない低グループ（送り仮名，仮名遣い，各種用語，流行語，情報化社会，日本語教育），１割～２割程度の中グループ（漢字，外来語の表記，語彙・用語，外来語，方言，言語政策，国語教育，外国語習得），３割～５割の高グループ（言葉遣い，あいさつ，呼称，敬語）とに分けられそうである。キーワード候補によって記事数は大きく異なるが，記事数が同程度の場合でも投書の割合には大きな差が生じている。例えば，各種用語，言葉遣い，言語政策は，それぞれが付与された件数が３～４千件台と同じくらいでも，投書の割合は，低グループ，高グループ，中グループと異なっている。

投書が読者の意見をある程度反映するものと考えれば，投書の割合が低，中，高グループに分けられることから，読者がことばのどの面に関心を持っているかということを探ることも可能ではないだろうか。そして，記事数や割合だけでなく，記事の内容を丁寧に見ることによって，日本人がことばあるいは日本語のどの面について，どのような価値観を持っているのか，それがどのように変わったのか，あるいは変わらなかったのかを知る手がかりを得ることができるように思う。

8.今後の課題

本報告では，『データベース』の概略を述べ，単純集計結果をもとに，ことばに対する関心のあり方を探る際の手がかりを示したが，今後『データベース』を利用する際の留意点を指摘しておく。まず，キーワード候補が対応する記事内容の範囲の広さあるいは狭さが集計に影響を与えている可能性がある。また，記事を収集する際の基準が明確に定まっていないために，年によって，あるいは担当者によって，収集するかしないかの判断にゆれが生じている。さらに，記事検索情報の付与にあたって，作業マニュアルの理解のしかたやデータ処理のしかたが作業に参加した複数の人の間でゆれている。記事検索情報の点検とさらなる整備は今後の課題である。

その他，『データベース』の一般への公開形態の検討，記事そのもののデータベース化の問題，著作権の問題など多くの点を克服しなければならないが，一つひとつ解決していくつもりである。

[1] 「朝日新聞見出しデータベース」は，朝日新聞の1945年から1995年までの50年の縮刷版について，記事検索見出し約308万件を収録したデータベースである。記事の内容によっては，見出しだけで検索が容易なものもあるが，ことばに関するデータを効率よく検索することは難しい。

[2] キーワード候補「語彙・用語」は分野としての語彙・用語とは異なる。『データベース』について設定したキーワード候補の83項目の中には，語彙・用語の他に，外来語，各種用語（専門用語），流行語等が独立して存在している。

[3] 井上・辻野(1992) 及び井上・池田・辻野(1994)において，辻野は文字・表記に関する記事の年推移を国語政策に関連させた考察や，常用漢字表に対する読者の意識のあり方について投書記事を資料として分析した結果について報告している。

[4] キーワード候補「情報化社会」は広い範囲を想起させるが，『データベース』では，タイプライター，ワープロ等の情報機器やソフトウエアの利用とその影響に関する内容の記事に付与している。

参考文献:

井上優・池田理恵子・辻野都喜江(1994)　「国語研究所所蔵新聞記事を利用した研究について（覚え書）」『国立国語研究所報告107研究報告集15』　141-163
井上優・辻野都喜江(1992)　「『国語関係新聞記事データベース』について（中間報告）」『国立国語研究所報告104研究報告集13』　165-193
尾崎喜光(1995)　「発話がもたらす対人効果の研究(1)―投書におけるメタコミュニケーション行動の分析―」『国立国語研究所報告110研究報告集16』　1-31
国立国語研究所編(～1989)　『国語年鑑』　秀英出版
田上稔(1992)　「方言復権の軌跡」『武庫川女子大学言語文化研究所年報』第４号　15-31