

CA1695

動向レビュー

データ分析による『カレントアウェアネス』レビュー

はじめに

前号をもって本誌『カレントアウェアネス』(以下、CA)は300号を数え、そして、1979年8月の創刊から30年が経過した。途中、頒布範囲の拡大(1989年6月に国立国会図書館(以下、NDL)外への頒布を開始)や刊行頻度の変更(2002年6月に月刊から季刊に変更)を経て、掲載記事を充実させながら「図書館に関する最新情報の速報による提供」を続けている⁽¹⁾。

前号では、企画・編集に関わった研究者らにより、自身の経験に基づく視点からCAの歴史が紹介されているが⁽²⁾⁽³⁾⁽⁴⁾、本稿では、「データから読み取れる」掲載記事の傾向と変化を明らかにすることを目的に、1989年6月(第118号)から2009年6月(第300号)までの約20年分の記事を対象として、それらの書誌データとテキストデータを分析する。

1. 分析対象

NDLのウェブサイト「カレントアウェアネス・ポータル」では、NDL外への頒布を開始して以降(1989年6月刊の第118号以降)の記事の全文が公開されている。本稿は、それらウェブ版の記事のデータを分析対象にする。基本的には、1989年6月以降は、冊子版もウェブ版も同じ記事が掲載されているが、ごく一部に違いがある。索引5件および1991年発行分の記事のうちの6件は、冊子版にのみ存在し、逆に、外国語による記事4件(英語2件と中国語2件、ただし、

それらの和訳は冊子版にも存在する)と参照文献リスト1件は、ウェブ版にのみ存在する。

CAを構成している主な記事は、(1)近年話題となっているテーマの中から注目すべき事例を1つを取り上げ、背景等の解説を加えながらわかりやすく説明する「一般記事」、(2)ある特定分野について最近の動向をレビューすることを主眼としている「動向レビュー」(2002年6月より)、そして(3)図書館情報学の研究に関する概況把握を目的として、特定テーマに関する最近数年間の研究論文をレビューする「研究文献レビュー」(2003年9月より)であり、CAナンバーが一連番号として付番されている。これらの記事1,067件に、「用語解説」(Tナンバーが付番されている)32件⁽⁵⁾と、小特集の冒頭の付言など、その他の記事(一連番号は持たない)29件を合わせた計1,128件について分析を行う。それらの記事の全文テキストデータに加え、NDLによって各記事に付与された件名と主題タグも分析に用いる。また、一部の分析においては、比較のため、より速報性が高いメールマガジン『カレントアウェアネス-E』(以下、CA-E)の記事948件(2002年9月の試行版から2009年6月までの全記事)のデータも併せて用いる⁽⁶⁾。

本稿と同様に、CA掲載記事の傾向を扱ったものに、橋詰の分析⁽⁷⁾がある。橋詰は、NDL外への頒布開始より前の時期も含めて、1979年8月から2003年9月までの記事を対象に、累積索引による調査に基づき主題の傾向をまとめている⁽⁸⁾。今回の分析と比較可能な部分に関しては、橋詰の分析結果を適宜参照することにする。

本稿が分析対象とする範囲について、CAの年ごとの掲載記事数と平均記事長を示したものが図1である。記事の長さ・分量は、本文の語数(正確に言えば、

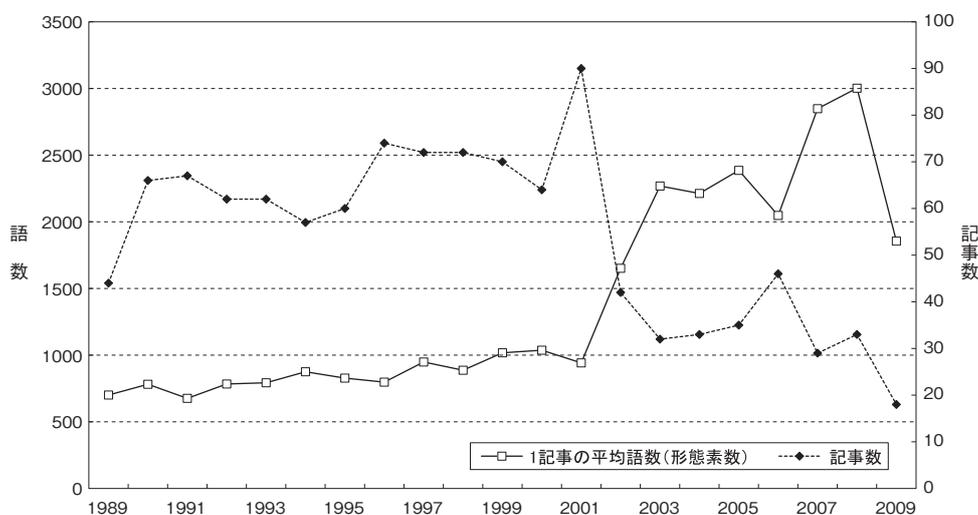


図1 記事数と1記事の平均語数の変化

単語より細かい単位である形態素の数)で計っている⁽⁹⁾。

2001年までは、記事の分量は緩やかに増加している。その後、2002年の季刊化を境に、年間の記事数が半減した一方で、1記事あたりの分量は倍増した。季刊化後に始められた「動向レビュー」と「研究文献レビュー」が、分量の増加につながっている。2000年代後半には、さらに分量が増えており⁽¹⁰⁾、詳細で長い解説記事が掲載される傾向を確認できる。

2. 参考文献から見る傾向と変化

CAの記事には、末尾に注とともに参考文献(引用文献、典拠文献)が挙げられている。参考文献の数は、記事の学術性の指標になり得るし、参考文献の新しさは速報性の指標になり得る。ここではそうした観点から分析した結果について報告する。ただし、使用したデータでは、注の中から個々の参考文献を正確に分割して取り出す自動処理は難しく、参考文献の書誌情報の形式が必ずしも統一されていないため、参考文献の出版年を完全に把握することも困難であった⁽¹¹⁾。また、末尾の注・参考文献リストではなく、本文中にのみ挙げられているCA、CA-Eの記事⁽¹²⁾は含めていない。したがって、図2・3に示す値は、大まかな傾向をつかむための概数である。

図2には、CAの1記事あたりの平均参考文献数とプライス指数を示した。プライス指数については後に述べる。図から、参考文献数は、(1)2000年まで緩やかに増加し、(2)2001-2003年にはやや大きく増加し、そして、(3)2004年以降急激に増加していることがわかる。(2)は、2002年に開始した「動向レビュー」、およびその前身である「Trend Review」(2001年10月～2002年1月掲載)の影響が窺える。(3)に関しては、2003年後半からの「研究文献レビュー」の影響が大きいと、それ以外の種別の記事も、

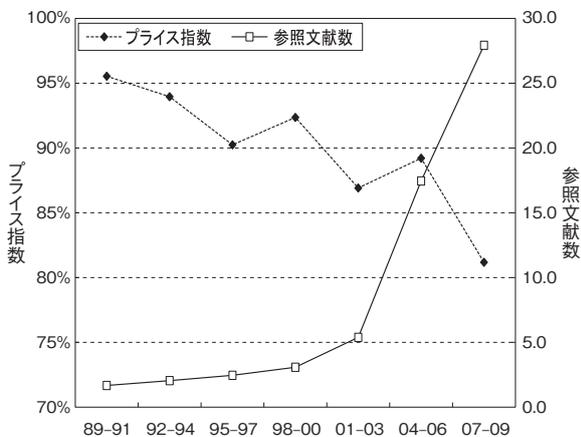


図2 参考文献の数とプライス指数

多くの文献を典拠として挙げるようになってきているという全般的な傾向がある。

次に、出版年を判定できた参考文献に関して、参照している記事と、参照されている文献の出版年の差、すなわち引用年齢を調べた。引用年齢の値が低いほど、新しい情報を参照していることを示す。参考文献として挙げられたウェブ情報源の中には、出版年の記述がなく、アクセス日付(最終確認日)だけが記されたものがあるが、アクセス日付では、情報の新しさがわからないので、それらのウェブ情報源は除いている。図3は、引用年齢ごとに、参考文献の割合を示したものである。CAで参照されている文献の半分以上が、概ね1年以内に出版された非常に新しい文献であることが確認できる。

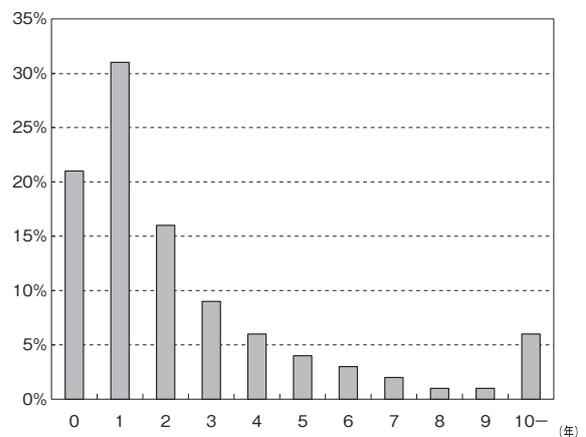


図3 引用年齢の分布

図2に参考文献数と併せて示したプライス指数も、参考文献の「鮮度」を表す指標である。プライス指数は、「参考文献の総数に対する、引用年齢が5年以内のもの割合」であり⁽¹³⁾、ここでは、記事単位の平均値でなく、当該時期における記事集合全体に対して求めた値を示している。

プライス指数は、2006年までは、徐々に低下しているものの、90%前後という高い値を保ちながら推移している。しかし、2007年以降になると大きく下降し、約80%まで落ちている。一般的には、プライス指数80%というのは、必ずしも低い値とは言えないが、CAが速報性を重視する雑誌であることを考慮すると「意外に低い」結果である。ただし、前述のとおり、出版年の記述がないウェブ情報源や、本文の中にだけ挙げられているCA、CA-Eの記事は含んでいないため、実質的なプライス指数は、もっと高くなると考えられる。また、図2に示されているように、プライス指数の低下が、参考文献数の増加とともに生じているものであることから、新しい情報をあまり参照しなくなったわけではなく、より以前の情報・

文脈も参照しながら新しい情報を紹介、解説するようになった、と解釈することができよう。この傾向は、本来「最近数年間」の論文が対象である「研究文献レビュー」でも見られる。

3. 扱われる内容の傾向と変化

CA で扱われている内容の傾向を、付与された件名と主題タグ、およびタイトル・本文中に出現する用語をもとに分析する。まず、統制語である件名と主題タグに基づき、いくつかの観点からおよその傾向を調べた後、出現用語のデータを分析して少し細かく見ていく。

3.1 件名、主題タグに基づく傾向

橋詰の分析にもある、(1) 国別（橋詰の分析では国内・国外別）の記事数と、(2) 館種別の記事数を調べた。国別の記事数は、国立国会図書館件名標目表 (NDLSH)、米国議会図書館件名標目表 (LCSH) による件名（一部フリーターム）に基づいて集計した結果を示す。主標目だけでなく細目としての付与も集計に含めている。館種別の記事数については、件名では「公共図書館」が1件も存在しなかったため、「カレントアウェアネス・ポータル」において表示される主題タグに基づいて集計した結果を示す。「公共図書館」以外は、件名と主題タグの集計結果の傾向にそれほど大きな差異はなかった。

記事数が多い4か国—日本、米国（件名表記は「アメリカ合衆国」）、英国（件名表記は「イギリス」）、中国—について、全体に占める割合の推移を示したのが図4である。米国は、2006年まではコンスタントに高く、概ね20%以上、つまりCAの記事5件に1件は米国を扱ったものだった。直近の2007年～2009年には、米国と英国の記事の比率が大きく落ち込んでいる。英米の比率が下がるのとは反対に、日本と

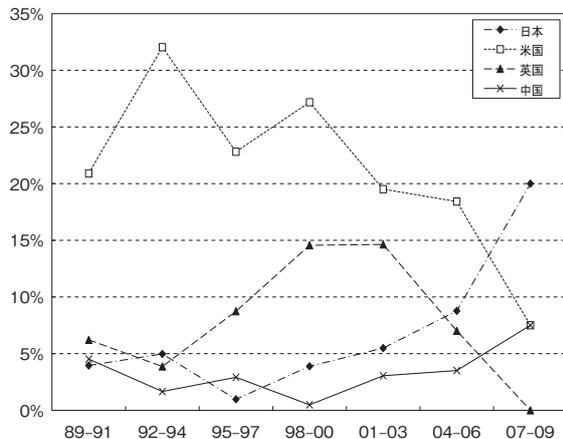


図4 国別記事の割合の推移

中国の比率は上がってきている。特に、日本の比率の上昇が大きい。橋詰の分析では、国内を扱った記事の割合は、2000年まで減少していることが報告されているが、本分析の結果から、その後、増加に転じていることがわかる。2003年に開始した「研究文献レビュー」が、国内の研究を対象にしていることも、日本の記事の割合が増加した一因であろう。

図5には、館種別記事の割合の推移を示した。大学図書館と研究図書館は1つにまとめている。図に表示のない学校図書館など他の館種は記事数が少なかった。2003年までの傾向は、橋詰の報告と類似している。すなわち、およその趨勢として、1994年までは全体的に割合が上がり、その後、NDL、国立図書館の割合は下がっている。2004年～2006年での落ち込みはあるが、公共図書館は高い割合を占め続けている。2007年～2009年になって国立図書館、公共図書館、大学・研究図書館が揃って大きく増加しているのは、納本制度や公共図書館の小特集が組まれたこと、機関リポジトリなど学術情報流通がホットなトピックになったことも理由として挙げられる。2004年以降、国立図書館、公共図書館、大学・研究図書館の比率はほぼ同程度であり、バランスがとれていると言える。

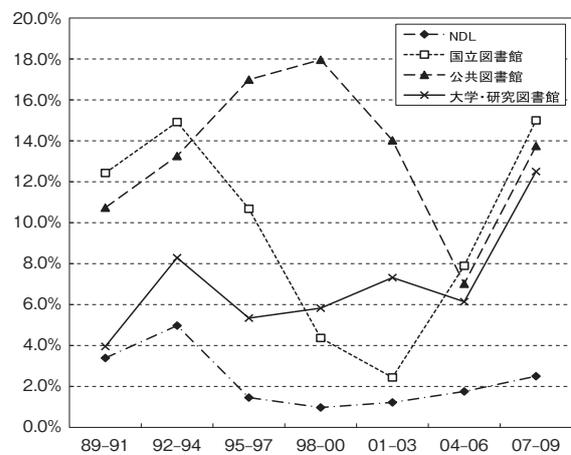


図5 館種別記事の割合の推移

3.2 出現用語に基づく傾向

タイトルと本文に現れる用語をもとに、CA掲載記事の傾向を分析した結果を示す。前節の件名・主題タグの付与状況とは異なり、「どんな用語が含まれているか」に関する情報であって、扱われているテーマを直接表すものではないことに注意を払う必要があるが、細かいトピックまでつかむことができる。重要度が高い用語のうち⁽¹⁴⁾、記事の割合の増加量・増加率が大きいものを表に挙げた。7年区切りで、前期（1989年～1995年）、中期（1996年～2002年）、

後期（2003年～2009年）に分け、前期から中期への変化と中期から後期への変化とをそれぞれ示している。

表 記事比率の増加が大きい用語

前期から中期への変化

	記事 比率 89-95 (A)	記事 比率 96-02 (B)	(B)-(A)
インターネット	0.060	0.351	0.291
Web	0.002	0.155	0.153
電子化	0.050	0.136	0.086
電子情報	0.036	0.116	0.080
ホームページ	0.002	0.081	0.078
博物館	0.038	0.110	0.071
公共図書館	0.261	0.331	0.070
情報資源	0.031	0.099	0.068
電子ジャーナル	0.002	0.070	0.068
電子図書館	0.043	0.110	0.066
紙媒体	0.002	0.068	0.066
コンソーシアム	0.002	0.064	0.062
電子出版	0.036	0.097	0.061
電子媒体	0.010	0.066	0.057
レファレンスサービス	0.012	0.066	0.054

中期から後期への変化

	記事 比率 96-02 (A)	記事 比率 03-09 (B)	(B)-(A)
評価	0.231	0.455	0.224
検索	0.287	0.486	0.199
大学図書館	0.248	0.410	0.162
図書館サービス	0.163	0.324	0.161
電子ジャーナル	0.070	0.225	0.155
情報資源	0.099	0.252	0.153
データベース	0.293	0.446	0.153
デジタル化	0.035	0.185	0.150
リポジトリ	0.002	0.149	0.147
図書館利用	0.058	0.185	0.127
図書館情報学	0.079	0.198	0.120
著作権	0.130	0.248	0.118
研究成果	0.025	0.140	0.115
デジタル情報	0.012	0.126	0.114
学術情報	0.043	0.153	0.110

前期から中期にかけては（表上）、「インターネット」「電子ジャーナル」など、電子情報関係の記事比率の増加が目立つ。この点は、橋詰の報告でも触れられている。「紙媒体」の増加も大きいですが、これは「電子媒体」との対比で語られているものである。電子情報関係の中でも、特に「インターネット」の増加が顕著であり、中期になると3分の1以上の記事に「インターネット」が含まれる。

中期から後期にかけては（表下）、「電子ジャーナル」が引き続き伸びているほか、「大学図書館」「リポジトリ」「学術情報」など、大学・研究図書館と関係が強い用語の伸びも大きく、前述の主題タグの分析結果と符合する。「リポジトリ」は、中期までは0%に近かった記事比率（2002年の1件のみ）が、後期になって15%近くまで伸びており、最初に世に出てから急速に広まったことが見て取れる。「評価」は、中期から後期にかけての記事比率の増加量が最も多い用語である。図書館のサービス／システム／職員／総合的なパフォーマンスの評価、資料・メディアの評価、情報の評価、図書館学教員の評価、研究評価といった、多様なトピックが取り上げられている。図書館界に限らず、近年、評価の社会的要請は強まっており、それが反映したものと考えられる。

「評価」という用語が、実際にどのような形で記事中に出現しているかを示すために、用語の変形を認識する関連語句検索システムを使用した⁽¹⁵⁾。図6は、中期と後期それぞれを対象に、「図書館評価」を検索語として記事のタイトル・本文を検索した結果である。記事から抽出した「図書館評価」の関連語句が出力されている。左列（検索語の下）は、検索語と同義の語句、および検索語と組み合わせられて現れている語句、右列は検索語の下位概念を表す語句である。後期（図6下）の方が、中期（図6上）に比べて、多様な形で現れていることがわかる。

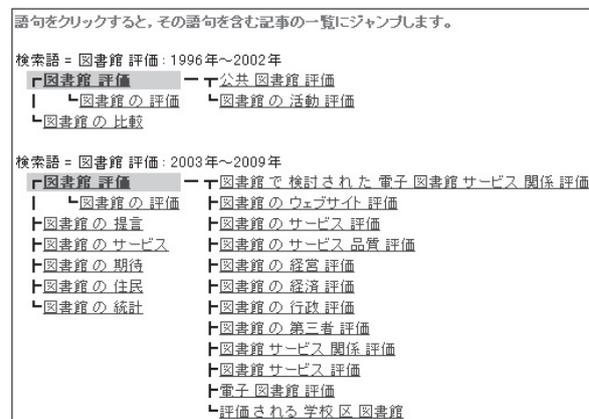


図6 「図書館評価」に関連する語句

表に挙げた各々の用語について、記事比率の推移をCAとCA-Eとで比較してみたところ、CA-Eの方が全般的に記事比率が低かったが、これは、本文が短く、含まれる語が少なく限定されていることによるものだろう。経年変化の挙動に関しては、だいたい似通っており、速報性の程度の差によるタイムラグは年単位では観察されない。一例として、CAとCA-E、それぞれの「リポジトリ」の記事比率を図7

に示した。「リポジトリ」は、先述のとおり 2002 年に、ちょうど CA-E 創刊と期を同じくして CA に現れ始めた用語である。

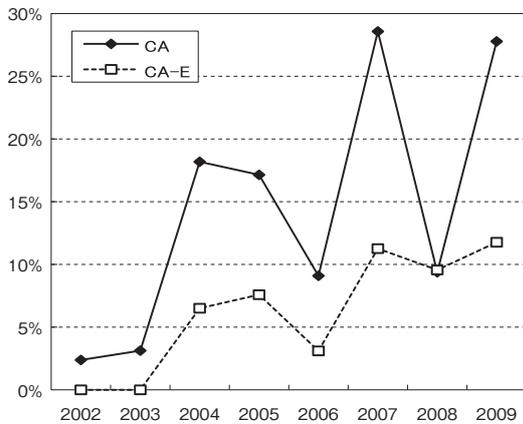


図7 「リポジトリ」の記事比率の推移

おわりに

その時々話題になっているテーマ、トピックを取り上げて解説する CA は、図書館界の動向を映す鏡と言える。CA の今後の方向性として、橋詰は、「インターネットによって“新鮮な”情報が容易に手に入るようになった現在こそ、レビュー誌が提供する“マクロな視点”はこれまで以上に不可欠となってくる」⁽¹⁶⁾と述べている。橋詰の指摘から 5 年が経過した今日、参考文献の分析で見たように、CA は、新鮮さと歴史的な文脈をおさえたマクロな視点とを兼ね備えたレビュー誌として着実に成長していると感じる。今後のさらなる成長を期待したい。

よしかわふゆき
(筑波大学：芳鐘冬樹)

- ム『茶筌』version 2.2.1 使用説明書．奈良先端科学技術大学院大学松本研究室，2000，21p.
- (10) 2009 年の記事数が少ないのは、6 月までの半年分のデータのみを対象にしているためである。そして、それら 18 件のうちの半数近くの 8 件が創刊 300 号・30 周年記念関係の特殊な記事(巻頭言など)であるという事情により、前年や前々年と比べて、2009 年の平均記事長は短くなっている。
 - (11) 書誌事項の記載順、および出版年の記載形式に関するいくつかのパターンを想定した正規表現によるマッチングで、各参考文献の出版年を判定した。
 - (12) CA は CA ナンバー(用語解説は T ナンバー)で、CA-E は E ナンバー(試行版は S ナンバー)によって参照指示が出されている。
 - (13) プライス指数は、参照・引用されている側の文献や雑誌などに対し「全被引用件数に占める、5 年以内に引用された件数の割合」として定義されるが、ここでは、参照・引用している側である CA の記事から見たプライス指数を算出している。
 - (14) 用語の抽出と重要度の計算には、中川らの接続頻度に基づく手法を使っている。「情報」のような一般的な語(主に単名詞)は除いている。
中川裕志，森辰則，湯本紘彰，出現頻度と接続頻度に基づく専門用語抽出．自然言語処理，2003，10(1)，p. 27-45.
 - (15) 用語の形態統語的変形をルールベースで認識する、芳鐘らの文書検索システムを使用した。
Yoshikane, Fuyuki; Tsuji, Keita; Kageura, Kyo; Jacquemin, Christian. Morpho-syntactic rules for detecting Japanese term variation: Establishment and evaluation. Journal of Natural Language Processing. 2003, 10(4), p. 3-32.
芳鐘冬樹，井田正明，野澤孝之ほか．キーワードの関連用語を考慮したシラバス検索システムの構築．日本知能情報フレンジ学会誌．2006，18(2)，p. 299-309.
 - (16) 橋詰秋子．動向レビュー誌『カレントアウェアネス』の役割と新たな展開．情報の科学と技術．2004，54(3)，p. 124.

- (1) 国立国会図書館関西館図書館協力課．『カレントアウェアネス』30 年の歩み．カレントアウェアネス．2009，(300)，p. 3-4.
- (2) 田村俊作．『カレントアウェアネス』の編集に係わって．カレントアウェアネス．2009，(300)，p. 5-6.
- (3) 野末俊比古．『カレントアウェアネス』:「変わったこと」と「変わらないこと」．カレントアウェアネス．2009，(300)，p. 7-8.
- (4) 北克一．『カレントアウェアネス』300 号への道程．カレントアウェアネス．2009，(300)，p. 8-9.
- (5) 「用語解説」は、同号掲載の一般記事で用いられている用語に関して補足説明する記事であり、1995 年から 2001 年までに「グーテンベルグ計画」など 33 の用語 (T1 ~ T33) が取り上げられた (T31 と T32 は 1 つの記事にまとめられているため、記事数は計 32 件である)。
- (6) CA、CA-E ともに、NDL から提供を受けた XML 形式のファイルを使用した。
- (7) 橋詰秋子．動向レビュー誌『カレントアウェアネス』の役割と新たな展開．情報の科学と技術．2004，54(3)，p. 120-125.
- (8) 本分析では、NDL 外への頒布開始前の記事は、著作権上の制限があり使用できなかった。
- (9) 本文の分析では、タイトル、著者名などの書誌事項、および注・参考文献は除いている。図表は、キャプションのみを含めている。また、外国語で書かれた記事 4 件は除外した。なお、テキストデータの形態素への分割と品詞の判別には「茶筌」を使用した。以後の分析でも同様である。
松本裕治，北内啓，山下達雄ほか．日本語形態素解析システ