

# 数理的研究

長谷川 守 寿

## 1. はじめに

本稿では、日本語に関連する主要学会誌に掲載された論文、さらに「NII 論文情報ナビゲータ CiNii」や国立国語研究所「日本語研究・日本語教育文献データベース」等から検索可能な論文を中心として、今期の数理的研究について整理する。また、学会の研究大会、ワークショップの予稿集に収録された研究も適宜紹介する。数理的研究であっても他分野に重点を置く論考については除外する。

紙幅の制限により、割愛せざるを得なかった文献が多々あるため、記述に偏りが生じたり、不十分な点や誤解もあるかと思う。筆者の力量不足を前もってお詫びしたい。

## 2. 動向

今期の大きなトピックは、日本語に関する初の書き言葉均衡コーパスを目指し2006年4月より国立国語研究所を中心として構築が進められてきた「現代日本語書き言葉均衡コーパス」(Balanced Corpus of Contemporary Written Japanese: 幾つかのバージョンが存在するが、以後全てBCCWJとする)が完成し、検索サイトとDVDによる全データが公開されたことであろう。完成に至るまでにモニター公開版が提供され、「日本語話し言葉コーパス」とあわせ、コーパスがより身近に使えるようになり、これらを用いた研究が盛んに行われるようになった時期とも言える。それを表すように学会誌では『言語研究』139号(2011.3)で「コーパスを活用した言語研究」という特集も組まれている。また解説書や論文集も幾つか出版されている。以下、研究対象別に概観していく。

## 3. 特定の表現に関する研究

大量の実例から意味・機能を分類し、表現の特徴を明らかにする研究は、コーパスを使って初めて可能になるものである。例えば、「～を秘書に雇う」の「秘書に」のような助詞「に」を伴い、役割を表す成分を対象とした(1)丸山直子「助詞「に」を伴う〈役割〉成分——コーパスに基づく分析——」(『日本語文法』10-1, 2010.3)や、「触る」が二格/ヲ格をとる場合の要因を考察した(2)秋元美晴「コーパスに基づいた「触る」の分析——他動性との関連から——」(『恵泉女学園大学紀要』22, 2010.2)等がある。また接続詞の研究とし

て、「また」「そして」「さらに」等の体系を整理した(3)中俣尚己「並列を表す接続詞の体系的分析」(『日本語文法』10-1, 2010.3)が挙げられる。

動詞に関しては、意味論的には自動詞用法しか持たないとされる「輩出する」「発生する」等の他動詞用法を分析した(4)小柳昇「コーパスに基づいた漢語サ変動詞の他動詞用法の分析——「場主語構文」の観点から——」(『言語・地域文化研究』16, 2010.3)や、自動詞・他動詞が同形である「一化する」動詞における自動詞用法・他動詞用法・「～させる」形使用・「～される」形使用について考察した(5)木山幸子・玉岡賀津雄「自他両用の「一化する」における自動詞用法と他動詞用法の比較——新聞コーパスの用例に基づく多変量解析——」(『言語研究』139, 2011.3)等が見られる。

ある表現の持つ意味・機能を対象とした研究として、否定形を含み目的を表すタメニ節について先行研究を検証し実態を明らかにした(6)茂木俊伸「目的を表す「ないために」の実態」(『鳴門教育大学研究紀要』25, 2010.3)や、「質問」・「疑い」という機能対立が認められる「のではないか」を分析した(7)佐藤雄亮「「のではないか」における〔質問〕と〔疑い〕の差異——BCCWJの用例分析から——」(『日本語文法』10-2, 2010.9)が挙げられる。また、ある表現の略語や表記のゆれに関する研究として、「リストラクチャリング」と「リストラ」等4事例について報告した(8)クドヤーロワ・タチアーナ「現代新聞における略語使用の変動傾向」(『計量国語学』28-3, 2011.12)や、「ゼリー」と「ジェリー」のような外来語のカタカナ表記のゆれが持つ機能を確認した(9)吉田充良「外来語のカタカナ表記のゆれと意味」(日本大学『語文』138, 2010.12)がある。

#### 4. 類義表現に関する研究

類義表現の研究では、(10)新屋映子「類義語「状況」「状態」の統語的分析——コーパスによる数量的比較——」(『計量国語学』27-5, 2010.6)や、複合動詞「～回す」「～回る」「～尽きる」「～果たす」等それぞれを考察対象とする杉村氏の一連の研究(例えば、(11)杉村泰「コーパスを利用した複合動詞「一回る」の意味分析」(『言語文化論集』32-2, 2011.3))をはじめ、今期は多くの論考が見られた。

その中で通時的観点を持つものに、現代語におけるガ・ケレド類の機能分析を行った(12)宮内佐夜香「通時の変化を背景とした接続助詞ガとケレド類の機能についての調査——『現代日本語書き言葉均衡コーパス』を資料として——」(『都大論究』47, 2010.6)や、可能表現形式(可能動詞・レル(ラレル)・コトガデキル)の動向を分析した(13)李慶實「近代日本語の可能表現形式の推移について——明治以降の文学作品を資料として——」(『計量国語学』28-2, 2011.9)がある。また、発話年代と発話者の出生年代という変数に注目して、人を主語とする存在文での「ある/いる」の選択傾向、副詞「全然」「全く」と否定辞との共起傾向の強さ等の分析を行った(14)服部匡「話者の出生年代と発話時期に基づく言語変化の研究——国会会議録を利用して——」(『計量国語学』28-2, 2011.9)や、二字漢語を語基とする

「～性」「～率」のような名詞と、「大きい／多い」等の形容詞の共起傾向の推移を調査した(15)服部匡「程度の側面を持つ名詞とそれを量る形容詞類との共起関係——通時的研究——」(『言語研究』140, 2011.9)等は、発話者と発話年代に関する詳細な情報を持つ国会会議録の利点を生かした研究である。

類義表現の共起情報に関する研究では、オノマトベと動詞の共起パターンの多様性と偏重性を、レジスターやオノマトベの多義性等の観点から考察した(16)玉岡賀津雄・木山幸子・宮岡弥生「新聞と小説のコーパスにおけるオノマトベと動詞の共起パターン」(『言語研究』139, 2011.3)や、「～テモ～テモ」について共起する動詞の特徴を分析した(17)清水由貴子「[A テモ B テモ] 文の分析」(『日本語文法』10-1, 2010.3)、(18)中溝朋子・坂井美恵子・金森由美・大岩幸太郎「漢語名詞「進歩」と「向上」のコロケーションの異同について」(山口大学大学教育機構『大学教育』8, 2011.3)等が挙げられる。また「～出す」を持つ複合動詞について、「～ダス」の部分の表記、前接する動詞のバリエーション、ジャンルによる違い等を分析した(19)石川慎一郎「現代日本語書き言葉均衡コーパス(BCCWJ)における複合動詞「～出す」の量的分析」(『統計数理研究所研究レポート』238, 2010.3)は、共起情報や表記、位相等、複数の観点を併せ持つ研究例である。日本語教育では共起情報も含めた語彙教材が出版されはじめており、この分野での研究成果は益々重要になってくるであろう。

## 5. 文章・文体に関する研究

語彙の量的分布を把握する新しい指標としてh指標を用いた研究に、(20)鈴木崇史・富坂亮太・内山清子・相澤彰子「h指標を用いたテキストの特徴分析」(『言語処理学会第16回年次大会発表論文集』, 2010.3)や(21)真田治子「h指標を用いた語彙の計量的分析」(立正大学『経済学季報』60-3・4, 2011.3)があり、今後のテキスト分析への応用が期待される。さらに「日本語教育」について言及した会議と人と論点を調査した(22)山本冴里「国会における日本語教育関係議論のアクターと論点——国会会議録の計量テキスト分析からの概観——」(『日本語教育』149, 2011.8)は、特定の表現からのテキスト分析であるが、使用したKH Corderというツールの可能性としても興味深い。

また文章ジャンルに関し、(23)中尾桂子「品詞構成率に基づくテキスト分析の可能性——メール自己紹介文、小説、作文、名大コーパスの比較から——」(『大妻女子大学紀要 文系』42, 2010.3)ではメール自己紹介文の特異性を明らかにし、(24)村田年・山崎誠「手」の慣用句を指標とした文章ジャンルの判別——現代日本語書き言葉均衡コーパスを用いて——(慶應義塾大学『日本語と日本語教育』39, 2011.3)では「人文科学系」「社会科学系」「自然科学系」というジャンル判別を、慣用句を用いて試みている。

文学作品・批評の計量的分析としては、村井・往住氏や村上氏のグループの研究が挙げられる。例えば、(25)工藤彰・村井源・往住彰文「計量分析による村上春樹長篇の関

係性と歴史的変遷」(『情報知識学会誌』21-1, 2011.2)では、12作品を対象とし、品詞分類の4品詞と意味分類の17カテゴリから分析を行い、(26)上阪彩香・村上征勝「西鶴作品の文章分析——先行研究の計量文献学的検証——」(『人文科学とコンピュータ研究会報告』2011-CH-90-2, 2011.5)では、西鶴作とされる作品の品詞の出現率と助動詞の出現率を主成分分析で分析し、先行研究の検証を行っている。また、読点が付く文字という特定の表現の頻度から夏目漱石の10作品を分類した(27)大久保起延・大久保博美「読点の施し方と漱石作品の分類」(『計量国語学』28-1, 2011.6)のような研究も見られた。

## 6. 辞典記述に関連する研究

従来、辞典の記述は編集者の知識や内省によるところが多かったが、BCCWJのようなコーパスの出現は今後大きな変革をもたらす可能性がある。(28)丸山直子「動詞の格情報——国語辞典の記述とコーパス——」(『日本文学』107, 2011.3)では、コーパスから得られた情報をどのように反映するか検討し、実際の例文の見直し作業で得られた知見を述べている。また「古語的」「古風」等の位相情報を持つ語の使用状況を調査した(29)柏野和佳子・奥村学「国語辞典に「古い」と注記される語の現代書き言葉における使用傾向の調査」(『人文科学とコンピュータ研究会報告』2010-CH-88-6, 2010.10)や、「卑俗体」「口頭体」「文語体」といった語の位相情報を明確にする方法を提案した(30)井上次夫「コーパスに基づく「語の文体」の明確化」(『白鷗大学教育学部論集』4-1, 2010.4)は、今後辞典の記述を充実させる有用な情報をもたらすと思われる。

なお(31)田野村忠温「日本語コーパスとコロケーション——辞書記述への応用の可能性——」(『言語研究』138, 2010.9)は、辞典だけでなくコロケーション辞典の編集にコーパスをどう生かすかという応用的、実用的な問題意識に基づいて行われた考察である。

## 7. 特定コーパスの構築に関連する研究

(32)孫栄爽「擬音語・擬態語と身振り——テレビ放送のマルチメディア・コーパスによる計量的分析——」(『計量国語学』27-4, 2010.3)では、言語だけでなく使用場面の映像・音声も参照できる特性を生かし、オノマトペと身振りの関係等が論じられており、(33)坂本真樹「小学生の作文コーパスの収集とその応用の可能性」(『自然言語処理』17-5, 2010.10)では、オノマトペの学年別使用実態の推移が調査されている。

また、(34)橋本力・黒橋禎夫・河原大輔・新里圭司・永田昌明「構文・照応・評価情報つきブログコーパスの構築」(『自然言語処理』18-2, 2011.6)では、固有表現や評価表現に言及し、(35)関洋平・神門典子・稲垣陽一・栗山和子「新聞記事とコミュニティQAを対象とした詳細な意見分析コーパスの作成と分析」(『情報学基礎研究会報告』2010-FI-97-6, 2010.1)では、意味タイプの出現の違い等を明らかにしている。他には(36)山口真紀・菅谷有子・単娜・古市由美子・村田晶子「工学系話し言葉コーパスにおける

和語動詞の使用実態——名詞との共起パターンの調査——』（『専門日本語教育研究』12, 2010.12）等、様々なコーパスが構築されており、公開等も含め今後が楽しみである。

## 8. 教育的観点を持つ研究

今期は日本語母語話者と日本語学習者の使用する日本語の差異を数量的に明らかにし、教育に役立てようとする研究が盛んに行われた。例えば、(37)中西久美子「日本語学習者・日本語母語話者のとりたて助詞の使用実態」（『計量国語学』27-7, 2010.12）、(38) 朴仙花「OPI データにみる日本語学習者と日本語母語話者による文末表現の使用——接続助詞で終わる言いさし表現を中心に——」（『言葉と文化』11, 2010.3）、(39)伊集院郁子・高橋圭子「日本語の意見文に用いられる文末のモダリティ——日本・中国・韓国語母語話者の比較——」（『東京外国語大学留学生日本語教育センター論集』36, 2010.3）、(40)耿梅晶「中国語を母語とする日本語話者による「から」「ので」の使用状況の考察——日本語母語話者との比較を通して——」（『日本語／日本語教育研究』2, 2011.5）等である。また(41)森篤嗣・庵功雄編『日本語教育文法のための多様なアプローチ』（ひつじ書房, 2011.10）にはコーパスを用いた様々な論文が解説と共に収められている。学習者コーパスの整備・公開も進み、この分野はさらに多様な研究が期待される。

さらに、教材に関連し、(42)李在鎬「大規模テストの読解問題作成過程へのコーパス利用の可能性」（『日本語教育』148, 2011.4）は、読解問題作成での使用を想定し、日本語能力試験の級区分に基づき BCCWJ の分析を行っている。また、小中高 12 学年を難易尺度として日本語テキストにおけるリーダビリティ判定式を構築した(43)柴崎秀子・原信一郎「12 学年を難易尺度とする日本語リーダビリティ判定式」（『計量国語学』27-6, 2010.9）も関連する研究として挙げられる。

## 9. 解説書

コーパスやウェブ等も含めた電子化テキストや形態素解析ソフトの普及にともない、計量的な研究を行うことに対する垣根は低くなっていると思われる。そのような希望を持つ人のためには、(44)荻野綱男・田野村忠温編『講座 IT と日本語研究』（明治書院）が出版されている。例えば、第 6 巻『コーパスとしてのウェブ』（2011.7）では、検索エンジンの特徴から方法、結果を利用した研究例等、詳細な情報が得られる。

また、(45)前川喜久雄「コーパスと言語学」『シリーズ朝倉〈言語の可能性〉6 言語と情報科学』（朝倉書店, 2011.7）は、言語学研究におけるコーパスの意義や種類、及び構築に関する問題等を述べている。さらに、数理的研究で重要な位置を占める統計学の視点や手法について詳細に解説した(46)石川慎一郎・前田忠彦・山崎誠編『言語研究のための統計入門』（くろしお出版, 2010.12）や、実際にコーパスを統計的な手法で分析し解説を加えた(47)李在鎬『コーパス分析に基づく認知言語学的構文研究』（ひつじ書房,

2011.2) 等が出版されている。

## 10. 最後に

今期使用されることが多かった BCCWJ であるが、今後益々数理的研究の対象として共有されることが予想される。但し、BCCWJ は出版・流通において均衡をとっているため、出版・流通年と執筆時期が異なるものもある。例えば福澤諭吉『学問のすすめ』は 2002 年にも出版されているため BCCWJ に含まれている ((48)丸山岳彦「大規模コーパスの利用とメタデータの役割」『第 1 回コーパス日本語学ワークショップ予稿集』(国立国語研究所, 2012.3))。これは、BCCWJ では明治以降を現代としているためであるが、研究目的によっては、このようなデータを他の書籍や白書、Yahoo! 知恵袋等と同等に扱うのか、研究を始める際に考慮すべきであり、コーパスの内容に関する知識の共有も重要になってくるであろう。

そこで最後に、電子化されたデータの本文批評として(49)荻野綱男「コーパスに含まれる「間違い」をめぐって(2)——BCCWJ と WWW の比較——」(『コーパスを利用した国語辞典編集法の研究』科学研究費報告書, 2010.2) や、『CD-HIASK'93 朝日新聞記事データベース』と新聞紙面を比較し、主に漢字について言及した(50)横山詔一「コーパス本文批評と統計的検定の考え方」(『講座 IT と日本語研究 5 コーパスの作成と活用』, 明治書院, 2011.6)、『CD- 毎日新聞』と新聞縮刷版を使用し、異同を論じた(51)長谷川守寿「新聞紙面と新聞記事データ集の相異について」(首都大学東京『人文学報』443, 2011.3) を挙げる。コーパスを用いた数理的研究を行う前に、その特徴をよく知ることが大切であると考え。また(52)李在鎬「コーパス基盤の言語研究」(『お茶の水女子大学比較日本学教育研究センター研究年報』6, 2010.3) は、望ましい調査を行う上で参考になる。

最後に、資料探しの際に、国立国語研究所研究図書室の加藤論子氏には大変お世話になった。記して感謝申し上げる。

——首都大学東京准教授——