

言語学の分野を中心に

森山 新

言語学の分野を中心に

森 山 新

要 旨

本稿では言語学の中から、近年その使用価値が注目され、発展のめざましいコーパスを用いた第二言語としての日本語習得・教育研究から4つを取り上げて概観した。本大会では日本語母語話者コーパスを用い、日本語教育の教科書、辞書、試験問題作成のための基礎研究を行った研究と、日本語学習者コーパスを用い、習得メカニズムを明らかにし、日本語教育に資することをめざす研究とがあった。近年の日本語コーパスの編纂により、これまでの設問調査、実験などによる引き出されたオンラインデータに加え、自発的なオフラインデータが大量に確保されるようになった。またコーパスの用い方も、単なる「コーパスで説明する研究」と、「コーパスを基盤とする研究」とがある。本稿で取り上げた4つの研究は後者になるが、コーパスの使い方をより詳細に見ていくと、「記述的、調査的に用いた研究」と、「説明的、仮説検証的に用いた研究」とがあった。コーパスの開発はこれまでの手法を補完するのみならず、精度を高め、日本語習得・教育研究により質の高い実証的根拠を提供してくれる。しかし日本語コーパスの開発は欧米の言語と比べまだ途についたばかりであり、規模や種類の面で限界があり、コーパスの「代表性」に問題を残してしまう場合が少なくない。さらに、日本語習得・教育研究にコーパスをどのように用いるべきかについても研究者の理解はまだ決して十分とは言えず、今後さらなる発展が期待される。

【キーワード】 コーパス (corpus),
コーパスで説明する研究 (corpus-illustrated research),
コーパスを基盤とする研究 (corpus-based research),
代表性 (representativeness)

1. はじめに

「言語学」は従来、人間言語の科学的研究をめざし、客観的な実験や観察などに基づく経験的な手法 (empirical method) をとってきた。しかし 20 世紀中盤に現れたチョムスキー (N. Chomsky) は、「言語能力 (competence)」と「言語運用 (performance)」とを区別し、言語学が取り組むべきは「言語能力」であるとした。その結果、言語学研究は一時、言語運用を重視する経験主義 (empiricism) 的なものから、生得的な言語能力を重視する合理主義 (rationalism) 的なものへと傾倒していった。同様に言語分析の手法もまた実験や観察といった客観的で経験的な手法でなく、母語話者の言語直観に依拠する主観的で内省的 (introspective) な手法が用いられるようになった。このような言語学全般の傾向は、日本語学や日本語教育学、第二言語としての日本語習得研究にも一時、少なから

ず影響を及ぼした。

しかしながら近年、第二言語としての日本語習得・教育研究の分野に新たな動きが感じられる。その1つが「コーパス言語学 (Corpus Linguistics)」の台頭によるコーパスデータの積極的な活用である。その結果、2003年には『日本語学』(明治書院)が4月号の臨時増刊として「コーパス言語学」と題した特集を組んだのを皮切りに、2004年には第二言語習得研究会が「電子化コーパスを利用した言語習得研究」をテーマにパネルディスカッションを行っている。2006年には『日本語教育』130号で「〔特集〕コーパスと日本語教育—現状と課題—」が、2007年に入ると『月刊言語』(大修館書店)7月号で「インターネットと言語研究—情報を選び分け、活用するために—」、さらに『日本語科学』22号では「コーパス日本語学の射程」という特集がそれぞれ組まれた。このようなコーパス志向の傾向は今回の国際研究大会の言語学関連の発表でも顕著に見られた。本稿ではこの点を中心に、関連の研究発表をふり返りながら、シドニーの大会を総括してみたいと思う。

上述のように、チョムスキーの合理主義的な言語学の考え方では、言語運用能力よりも生得的な言語能力に、また、具体的な言語事例よりも抽象的な言語知識に関心が向けられた。その結果、実験や観察、調査といった方法よりも、内省による研究がさかんに行われた。

これに対し日本語教育学を始めとする第二言語・外国語教育学では、生得的言語能力よりも、実際に学習者は何ができるかという言語運用能力に関心があり、この点ではチョムスキーの合理主義的な考え方とは対照的で、言語使用(運用)などの経験を重視する。

また第二言語習得研究の歴史をふり返ると、「教師がいかにかえるか」という教育的側面に関心を持っていた20世紀前半に対し、20世紀後半にはその前提として「学習者がいかに学ぶか」という習得的側面に対し注目が集まり、誤用研究、さらには中間言語研究などがさかんになった。それらでは学習者の実験や観察、設問調査などを通じてデータを集め、習得のプロセスやメカニズムを客観的に分析するという経験的手法がとられた。

今日コンピュータとインターネットが飛躍的な普及と発展をなしているが、こうした変化は言語学の分野にも影響を及ぼしている。データの電子化を飛躍的に発展させ、大規模なデータの収集と保存、アクセスと解析を可能とした。このようなデータはコーパス(corpus)と呼ばれ、言語使用のデータの貯蔵庫として、英語を始め、様々な言語で開発、利用できるようになった。

実際の使用(運用)を重視する第二言語習得・教育研究がこれを放っておくはずがなかった。英語教育などではかなり前からコーパスを用いた習得研究や辞書などの教材開発が行われてきた。しかしながら、日本語教育でのコーパス利用に関しては、これまで日本語コーパスの開発・整備の立ち遅れが指摘され、習得研究や教材開発に十分生かすことができなかった。

ところが最近、国立国語研究所が話し言葉、書き言葉のコーパスを開発するなど、様々なコーパスが登場している。モノログを中心とする現代日本語の自然な話し言葉コーパスである「日本語話し言葉コーパス (Corpus of Spontaneous Japanese: CSJ)」は662時間、およそ752万語分の音声記録されており、音声を提供している話し言葉コーパスとしては世界最大規模のコーパスで、2004年に第1刷が2008年には第2刷が公開されてい

る。また書き言葉コーパスでは「現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Japanese: BCCWJ)」の構築が2006年度から5年計画で進行中で、出版実態サブコーパス (新聞, 雑誌, 書籍), 流通実態サブコーパス (書籍), 非母集団サブコーパス (白書, 広報誌, 法律, 教科書, 議事録, ベストセラー, Web など) からなり、日本語ではこれまでにない1億語超規模の大規模均衡コーパスとなる予定である。現在、モニター用のデータが公開されている。「コーパス日本語学のための情報館」(「参考サイト」参照)によれば、現在代表的な日本語コーパスは、書き言葉が9, 話し言葉が6, ウェブデータが3, 言語習得が2, 辞書・シソーラスが6と26を数えるまでになっており、日本語コーパス開発も急成長を遂げている。こうしたコーパス開発を背景に、日本語習得・教育の分野の研究も始まっている。

2. コーパスを用いた日本語習得・教育研究

今大会でコーパスを用いた言語学関連の研究発表には以下のようなものがある (予稿集から「コーパス」で検索したもので、発表者の都合から当日発表が行われなかった (12) も含まれている。カッコ内は用いられているコーパスである)。

- (1) 本田ゆかり「大規模コーパスを用いた学習者語彙表の作成の試み」(BCCWJ)
- (2) 堀恵子「日本語能力試験文法出題基準の機能語を対象としたコーパス調査」(朝日新聞, 新書, 文系理系学術論文, CSJ学会講演・模擬講演・課題指向対話・模擬講演インタビュー・自由対話・学会講演インタビュー・BTSJによる日本語話し言葉コーパス, 上村コーパス, CallHome)
- (3) 土屋智行, 李在鎬, 横森大輔「日本語における形容詞の活用と使用頻度」(BCCWJ, 読売新聞)
- (4) 中俣尚己「程度増大の接続詞「さらに」「しかも」「そのうえ」の使い分け—何が增大しているのか」(CSJ, 毎日新聞 2002, 日本語学習者による日本語作文とその母語訳との対訳データベース)
- (5) 原田幸一「「ばい」の前接部分を伴わない用法」(goo ブログ検索)
- (6) 峯布由紀「文末表現の習得順序を決定する要因についての考察」(独自のデータ, KYコーパス)
- (7) 江田すみれ「会話と科学入門書における「ている」の用法の違い」(男性のことは職場編, 女性のことは職場編, 『CASTEL/J』所収の新書)
- (8) 宇佐美まゆみ「日本語母語話者の日本語コミュニケーションの実態—母語場面と接触場面の比較も交えて—」(BTSJによる日本語話し言葉コーパス)
- (9) 佐野洋「日本語教科書コーパスの構築」(自作日本語教科書コーパス)
- (10) 森山新, 荒川洋平, 今井新悟「認知言語学的観点からの日本語学習辞典を考える」(Sketch Engine)
- (11) 渡辺美知子「話者の性別・年齢が日本語のフィラー選択に与える影響」(CSJ 模擬講演)
- (12) 李在鎬「日本語の習得レベルの計算的推定について: 習得レベルを決定づける要因は

何か」(KY コーパス)

これによると、BCCJW を用いた研究は2件 ((1), (3)), CSJ が3件 ((2), (4), (11)) 見られる。この他に複数の研究で用いられているコーパスは、新聞記事データベース3件 ((2), (3), (4)), BTSJ による日本語話し言葉コーパス2件 ((2), (8)), KY コーパス2件 ((6), (12)) がある。

本稿ではこれらの研究のうち、日本語母語話者の言語使用コーパスを用いた(1), (2), (3), 日本語学習者の言語使用コーパスを用いた(6)について紹介しながら、コーパスを用いた日本語習得・教育研究の現状と課題を探る。

2.1 「大規模コーパスを用いた日本語学習語彙表の作成の試み」(本田ゆかり)

この研究は BCCWJ 領域内公開データ (2008 年度版) で公開されている4つのサブコーパス (書籍, 白書, 国会議事録, Yahoo!知恵袋) を使用し, 日本語学習語彙表 (以下「語彙表」) を作成することを目的としている。この語彙表はコーパスの頻度情報と統計指標 *D* (Juilland, et al. 1970) と, *DP* (Gries 2008) を用いて日本語の学習で学ぶべき語彙 (自立語) を客観的に抽出したあと, 学習の観点から補正する方法で作成, レベル別に提示することをめざしている。研究の結果, 以下のような点を明らかにしている。

- ① 品詞のバランス: 頻度順に 1000 語区切りにしたとき, 品詞のバランスはほとんど変化せず, 安定している。但し 1000 語レベルでは感動詞, 接続詞, 代名詞, 連体詞がわずかに高く, また接続詞, 代名詞は基本語レベルでほぼ全て出現するなど品詞のバランスに若干の差異も見出されている。
- ② カバー率: どのサブコーパスでも 3~4000 語ぐらいまではカバー率が上昇するが, それ以降は停滞する。
- ③ サブコーパスごとの相違: サブコーパスにより品詞の構成比はほぼ一致しているが, カバー率では書籍とその他で違いが大きく, 特に白書でカバー率が低い。サブコーパスにより語彙の分布に偏りがあるもののがかなりあった。
- ④ 日本語能力試験出題基準語彙との比較: 頻度上位 1000 語に 2 級, 1 級, 級外の語彙が 51% も含まれていた。言い換えれば高頻度の重要語彙が多数出題基準語彙に含まれていなかったり, 1, 2 級語彙となっていたりしていることがわかった。
- ⑤ 代表性: BCCWJ は個々のサブコーパスで構成比に特徴があり, 日本語学習語彙表作成には不向きで, 話し言葉的語彙が抽出できるようなサブコーパスを増やすなど, 日本語学習に適した構成比を考える必要がある。

①, ②は日本語教育, 特に語彙教育への応用可能性を含んでいる。また④は現在の日本語能力試験出題基準語彙の再検討に際し実証的な根拠となりうる。③, ⑤はさらなるコーパスの開発など, コーパスを用いた研究の課題を明らかにしている。

本研究で明らかになったように, 大規模コーパスの開発により, 経験的, 実証的手法から学習語彙表の検討が可能になり, より実用的な語彙表の作成への道が開かれた。その意

味で本研究のような研究が今後もさらに進められることが願われる。しかし、BCCWJは総合目的のため編纂された「汎用コーパス (general purpose corpus)」であり、学習語彙表作成といった特定の言語研究のために編纂された「特殊目的コーパス (special purpose corpus)」ではない。コーパス利用に際して重要なのは「代表性 (representativeness)」, すなわちコーパスが実際の言語使用をどの程度代表するサンプルとなっているか, である。それぞれの研究目的に合わせて, そのたびに大規模コーパスを作成するに越したことはないが, それは実質的に不可能なことである。その意味で「汎用コーパス」をある特殊目的のために用いるにあたり, どの程度代表性を保証できるのか, また, 代表性を保証するために, サブコーパスの構成比の検討など, どのような工夫が可能か, などの点が, 本研究も含め, 研究の質を高めるために重要な課題となるであろう。

2.2 「日本語能力試験文法出題基準の機能語を対象としたコーパス調査」(堀恵子)

この研究は, 今回の日本語能力試験 (JLPT) 改定で目標となっている「課題遂行のための言語コミュニケーション能力」測定のために, どのような文法項目を選定すべきかを明らかにするため, 出題基準1級「文法的な<機能語>の類」として挙げられている99項目が実際に使用されているのか, その使用頻度を調査している。用いられたコーパスは「日常生活・教育・職業の3場面」×「文章・口頭 (各々改まり度の高・低) の4カテゴリー」からサンプルとなるコーパスを選んでいる (但し今回の発表では調査未完了である「職業場面」「文章・改まり度低」は扱っていない)。具体的には文章表現としては朝日新聞32日分, 新書100万字分, 文系理系学術論文31篇, 口頭表現としてはCSJのうち首都圏に住む話者による学会講演, 模擬講演, 課題指向対話, 模擬講演インタビュー, 自由対話, 学会講演インタビュー, 「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)」による日本語話し言葉コーパス, 上村コーパス, 電話会話コーパスCallHomeの書き起こしデータが用いられた。

本研究が明らかにしているのは, 以下の点である。

- ① コーパスごとの出現頻度: 99項目の多くの項目で出現頻度が0または低い結果となった。これには今回扱ったコーパスの規模が小さく, たまたま現れなかった可能性, 及び今回のコーパスの代表性の問題 (出題基準を明らかにする目的と合致していたか) が原因として考察されている。
- ② 出題頻度から見る文法項目の特徴: クラスタ分析の結果, 頻度の高いものは多くの用例を持ち, 分野を越えて広く用いられる一般的概念を表す表現で, 頻度が低いものは新書, 模擬講演, 丁寧な対話, 論文など特定分野でしか現れないもの, 古語を含む硬い表現, 話者の主観的価値観を含蓄する表現などであった。
- ③ 出現頻度から見るコーパスの特徴: 7つのコーパスをクラスタ分析した結果, 口頭表現 (対話・学会講演・模擬講演) と新聞とで, また新書と論文とで, 出現頻度に類似性が見られた。

②, ③は文法項目には一般性の高いものと低いものがあることや文脈により必要とされ

る文法項目に違いがあることが具体的に示され、教育的応用の可能性を示している。①、③は本研究目的を達成するにあたってコーパス分析が抱える課題であり、これをどう克服していくかが今後の研究の鍵となるであろう。その一部はコーパス開発の課題でもある。

2.3 「日本語における形容詞の活用と使用頻度」(土屋智行, 李在鎬, 横森大輔)

本研究は、形容詞が活用形ごとにボトムアップ的に記憶され、習得されているのならば、そこに一定の意味的な違いが生じている可能性があるのではないかとの仮説のもと、「形容詞はその活用形によって意味の使用分布が異なるか」を明らかにしようとした研究である。例えば「快い」は、連体形では五感を通した快の感覚(ex.快い音色/風)、連用形では事態に対する歓迎の感情(ex.快く引き受ける/思う)と、活用形による意味の違いに一定の傾向が見られている。上記の2つの研究が主に使用頻度を明らかにすることで、語彙表や能力試験の開発といった日本語教育の応用を試みているのに対し、この研究では、クロス集計を行うことで、「活用形と意味との相関」(以下の④)などを調べ、明らかになった結果を辞書記述のために情報として提供しようとしているのが特徴で、その意味で同じコーパスに依拠した(corpus-based)研究でありながら、そのアプローチに進展が見られる。

BCCWJ、読売新聞11年分のデータがコーパスとして用いられている。データ抽出にはKHCorderを使用している。調査にはChaSenを用い、形態素解析の後、形容詞の終止形、連体形、連用形を抽出し分析を行った。全体として異なり語数389語、延べ語数23,150語のサンプルが抽出された。

明らかになったことは以下のような点である。

- ① コーパス間の活用形の分布には違いが見られず、一貫していた(0.80以上の相関)。
- ② コーパスごとに意味の使い分けに違いが見られ、書籍、新書では基本義が多く、国会議事録、新聞では拡張義が多くなっていた。
- ③ 語彙ごとに意味の使い分けに違いが見られ、「根強い、等しい、少ない」などは終止形、「深い、広い、楽しい」などは連体形、「根気強い、手早い、熱っぽい」などは連用形での使用が主である。また連体・連用修飾用法が主である形容詞のほうが多義的であった。
- ④ 意味は文脈、語彙との相関があり、文脈に依存し、語彙項目ごとに異なる。意味と活用形との全般的な相関(例えば、概して連体形は基本義で使われやすく、連用形は拡張義で使われやすいといった相関)は見られなかった。これは活用形とその意味の分布が語ごとに異なっていることが原因と考えられる。

このように本研究では、事前に立てられた仮説(意味と活用形との全般的な相関)は検証されなかったが、それは意味と活用形との分布が語彙項目ごとに異なっているからであることがわかり、その意味で新たな相関の発見につながっている。意味と活用形との相関もまた、ボトムアップであるがゆえの結果と言えるであろう。

また本研究は、形容詞が活用形別(項目依拠的)に、かつボトムアップ的に習得される

といった仮説を立てていることで明らかなように、使用基盤的な言語観に立っている。「使用基盤言語学 (Usage-based Linguistics)」はラネカー (R. Langacker) が提示した認知言語学 (Cognitive Linguistics) の言語 (習得) 観であるが、本稿で取り上げたコーパス言語学と同様、実際の言語使用を重視する立場を取っており、Usage-based という面で両者は密接な関わりを持っている。しかしながら、認知言語学を支持する多くの研究者たちはこれまで、具体的な言語使用を重視するという立場に立ちながらも、依然チョムスキー以来の「内省法 (introspection)」を安易に用い続けた結果、具体的な言語使用のデータであるコーパスを利用することは、最近まで決して多いとは言えなかった (Tummers et al. 2005, 早瀬・堀田 2005, 森山 2008)。本大会でも認知言語学的観点からの発表は数件見られたが、コーパスを利用している研究としては、この研究のほかに、(10)のみであった。しかし認知言語学とコーパス言語学とは近年急速に接近しつつあり、今後、認知言語学的観点からの理論や仮説に対し、コーパスデータを用いてその妥当性を検証していく研究が増えていくものと思われる。

2.4 「文末表現の習得順序を決定する要因についての考察」(峯布由紀)

この研究は、これまで峯が独自に収集したデータをもとに行っていた文末表現の習得順序の研究成果を、KY コーパスで確認し、さらに、順序決定に係わる要因として、①言語処理、②意味・機能の2つの面から考察したものである。

分析の手法は、3つのデータ (峯 (1995)、及び峯他 (2002)) で報告されている、(1)教室習得者 30名のデータ、(2)自然習得者4名のデータ、そして(3) KY コーパスのデータ) を用い、それぞれ学習者の日本語能力レベル別に、使用文末表現とその使用者数を集計した。そしてその分布から習得順序を判定し、ピーネマン (M. Pienemann) の「言語処理可能性理論 (Processability Theory)」の妥当性について考察を行っている。

そして、これらのデータ分析より明らかになったことは以下のような点である。

- ① 日本語のレベルが高くなるにつれて、使用される表現が増える順序に共通性が見られる。このことから、学習者が使用可能な表現は、環境が異なっても、日本語のレベルによる制約を受ける。
- ② 表現の使用に必要な言語処理に注目し、使用表現の広がり方を見てみると、語彙処理の階層で使用可能な表現は初級から使用されているが、複文処理を必要とする表現の使用は中級以上になってようやく使用されるといった傾向が見られる。これは「言語発達とは言語処理の発達であり、自動化が語彙・範疇処理→句処理→文処理→複文処理・文法処理と進む」とするピーネマンの「言語処理可能性理論 (Pienemann 1998)」を支持する結果と考えられる。
- ③ 意味・機能の面から使用表現の広がり方を見てみると、聞き手に働きかける機能を持つ表現 (発話の場面に密着した「今、ここ」の表現) の使用が早い。また、「一対一の原則 (One-to-one Principle)」により、先に使用される表現と似たような機能を持つ表現の使用は遅くなる傾向がある。
- ④ 習得順序は、「言語処理の階層」と「意味・機能」の両面からとらえるべきである。

①～④の結果はいずれも第二言語習得のメカニズムを示唆しており、日本語教育への応用可能性を感じさせる。通常、独自に収集されたデータが分析されている場合、学習者の日本語のレベルが初級、中級、上級と明示されていても、そのレベル判定が研究者や各研究で統一されていないため、他の言語項目との習得時期の比較が難しいという問題が生じる。したがって、上記①にあるように、データが異なっても、同じような習得順序が見られることを示したことはこの研究の成果と考えられるが、それ以上に、OPIという客観的な基準を持ったレベル判定により分けられた各レベルの学習者が、どのような表現を使用しているのかを具体的に示したことは、大きな成果であろう。

実際の日本語授業のシラバスやカリキュラム作成に習得順序の研究成果を取り入れることを考えるならば、文末表現の習得順序、テンス・アスペクトの習得順序といった、カテゴリ別の習得順序よりも、これらカテゴリを統合した習得順序を明らかにすることが必要であろう。なぜならば、文末表現、テンス・アスペクト、助詞、というようにカテゴリ別に授業をするわけではないからである。また、教授法の効果を見る研究でも、扱う言語項目はそのレベルで習得可能な言語項目であるかどうかで教授法効果の結果が変わりうる。習得順序通り教えるべきかといった議論はあるにしても、各レベルで習得可能な表現を研究し、提示していく基盤となる研究は必要不可欠であろう。

3. おわりに

本稿では言語学の中から、近年その使用価値が注目され、発展のめざましい「コーパス言語学」にフォーカスを絞って、コーパスを用いた第二言語としての日本語習得・教育研究から4つを取り上げて概観してきた。このうち2.1～2.3は日本語母語話者のコーパスデータを用いて、日本語教育の教科書、辞書、試験問題作成のための基礎研究を行ったものであり、2.4は日本語学習者のコーパスデータを用い、日本語の習得のメカニズムを明らかにすることで、日本語教育に資することをめざす研究である。近年の日本語コーパスの編纂により、これまでの設問調査、実験などによる引き出されたオンライン (on-line) データに加え、自発的なオフライン (off-line) データが大量に確保されるようになってきた。それぞれの研究手法はカバーできる研究分野や研究課題が異なっている。例えば実験データは言語処理のオンラインのプロセスを明らかにするのに向いている。これに対し、コーパスデータは言語産出の自発的なオフラインのデータであり、言語使用の実態を知るのには向いているが、言語処理のオンラインのプロセスを明らかにするのには不向きである。さらにコーパスの使い方も、単なる「コーパスで説明する研究 (corpus-illustrated research)」と、「コーパスを基盤とする研究 (corpus-based research)」とに分けることができる。後者はさらに「記述的、調査的に用いる段階 (descriptive and exploratory stage)」のものと、「説明的、仮説検証的に用いる段階 (explanatory and hypothesis-testing stage)」のものがある。前者は言語使用の注目する項目やその使用頻度について同定したり記述したりするもので、後者はその分布を説明したり理論構築を目指すものである (Tummers et al. 2005)。本稿で取り上げた研究はいずれも「コーパスを基盤とする研究」であると言えるが、このうち2.1、2.2は教材や試験の開発にあたり、語彙、文法項目の使用頻度について調べており、どちらかという記述的、調査的にコーパスを用いた研究で

ある。これに対し、2.3は意味と活用形の相関についての仮説を検証するもの、2.4は言語処理可能性理論の妥当性について検証したもので、両者はどちらかといえばコーパスを説明的、仮説検証的に用いた研究であると言えるだろう。

このようにコーパスの開発はこれまでの手法を補完するのみならず、精度を高め、日本語習得・教育研究により質の高い実証的根拠を提供してくれることが期待できる。しかし既に述べたように、日本語におけるコーパスの開発は近年めざましい発展が見られるというものの、英語など欧米の言語と比較してみればまだまだ途についたばかりであると言わざるをえない。そのためコーパスの規模や種類の面で限界があり、それが原因で研究目的との間にずれが存在し、コーパスの「代表性」に問題を残してしまう場合が少なくない。さらに、日本語習得・教育研究にコーパスをどのように用いるべきかについても、研究者の理解はまだ決して十分とは言えず、今後さらなる発展が期待される。

参考文献

- (1) 早瀬尚子, 堀田優子 (2005) 『認知文法の新展開－カテゴリー化と用法基盤モデル－』, 研究社.
- (2) 峯布由紀 (1995) 「日本語学習者における文末表現の習得過程に関する研究」『日本語教育』86号, 56-80.
- (3) 峯布由紀, 高橋薫, 黒滝真理子, 大島弥生 (2002) 「日本語文末表現の習得に関する一考察－自然習得者と教室学習者の事例をもとに」研究成果報告書『第二言語としての日本語の自然習得の可能性と限界 (平成 12-13 年度科学研究費 萌芽的研究 課題番号 12878043 研究代表: 長友和彦)』, 64-85.
- (4) 森山新 (2008) 『認知言語学から見た日本語格助詞の意味構造と習得－日本語教育に生かすために－』, ひつじ書房.
- (5) 吉野貴好 (2000) 「英語コーパス言語学の歴史的背景」『高崎経済大学論集』第 43 巻 第 1 号, 97-107.
- (6) Gries, S. (2008) Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics.*, 13-4: 403-437. John Benjamins.
- (7) Juilland, A., Brodin, D. and C. Davidovich (1970) *Frequency Dictionary of French Words*. The Hauge: Mouton
- (8) Pienemann, M. (1998) *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.
- (9) Tummers, J., Heylen, K. and D. Geeraertsm (2005) Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1 (2): 225-261.

参考サイト

「国立国語研究所 話し言葉コーパス」

<http://www.kokken.go.jp/katsudo/seika/corpus/>

「国立国語研究所の言語コーパス整備計画」

<http://www.kokken.go.jp/kotonoha/>

「コーパス日本語学のための情報館」 <http://www30.atwiki.jp/corpus-ling/>

「Sketch Engine」 <http://www.sketchengine.co.uk/>

(お茶の水女子大学大学院)

.....

The Presentations on Linguistics

MORIYAMA Shin

In this paper, I give an overview of four presentations of research in the acquisition and teaching of Japanese as a second language that make use of the corpus-based approach, which has demonstrated its utility in recent years with some remarkable developments. This includes the use of language corpora of native speakers in basic research to produce textbooks, dictionaries and test questions, and the use of corpora of second-language speakers to clarify the mechanism of language acquisition with the aim of supporting the Japanese Language Education. Thanks to the compilation of Japanese language corpora in recent years, a vast quantity of off-line linguistic data has been compiled, as well as on-line data elicited with surveys and experiments. The uses of corpora may be characterized as “corpus-illustrated research” and “corpus-based research.” The four studies discussed in this paper are the latter. In addition, a closer look reveals corpora using both descriptive and exploratory-based research, and explanatory and hypothesis-testing research. The development of corpora provides more high-quality empirical evidence, which enhances the precision of acquisition and teaching research and also complements previous approaches. Nevertheless, the development of Japanese-language corpora still lags behind developments in European- and American-languages research; they are limited in scope and variety, and there remain some problems with how representative they are. And, moreover, there seems to be a lack of understanding on the part of researchers of how corpora should be used for acquisition and teaching research. Therefore, it is to be hoped that progress will continue in the future.

(Graduate School of Ochanomizu University)

