

英語コーパス学会創立10周年記念第20回大会

日時 2002年10月5日(土)・6日(日)

会場 名古屋大学国際開発研究科(〒464-8601 名古屋市千種区不老町)

地下鉄東山線本山駅より徒歩20分、あるいは市営バス1番乗り場「本山」より島田住宅行、平針住宅行、名古屋大学前行に乗車、「名古屋大学前」下車、詳細は <http://www.gsid.nagoya-u.ac.jp/> 参照

大会テーマ わが国のコーパス言語学の成果と展望

10月5日(土)

ワークショップ(1) 10:30 - 12:00

《正規表現によるテキスト検索(基礎編)》

講師 名古屋大学 大名 力

定員 先着40名(予定) 参加費 会員無料・非会員は当日会費1,000円(申し込みは電子メール・郵便で事務局まで) なお、ワークショップ(1)の受付は2日目のワークショップ(2)の受付とは別に行っています。どちらを受講するかを明記してください。両方受講することも可能です。なお、当日は昼食をご持参下さい。

受付開始 12:30

開 会 13:00

1. 会長挨拶 摂南大学 今井 光規
2. JAECS 賞審査経過の発表・表彰
3. その他

研究発表 第1室

第1セッション 13:20 - 14:10

司 会 北海道大学 園田 勝英

1. アメリカ英語における国名の使用頻度に基づく「認知的世界地図」構築の試み
名古屋大学院生 松野 和子
2. アカデミックスピーカーの発話語数のジェンダー差
American-English の分析から
Corpus of Spoken Professional

摂南大学 家口美智子
京都大学 家人 葉子
神戸市外国語大学院生 岡部 浩子

第2セッション 14:20 - 15:10

司 会 名古屋学院大学 鈴木 重樹

3. 文の直接構成要素レベルでの解析済みコーパスの比較
ランカスター大学共同研究員 田中 泉
4. 文体差に関わる語彙の特定 Bank of English の「大衆紙語」と「高級紙語」の形容詞
北海道大学 高見 敏子

研究発表 第2室

第1セッション 13:20 - 14:10

司 会 明海大学 投野由紀夫

1. 国際学習者コーパス ICLE/LINDSEI 最新の動向
2. 英語学習者コーパス分析法の開発から電子掲示板の改良へ
昭和女子大学 金子 朝子
長崎純心大学 鈴木千鶴子
" 吉原 将太
" 渡辺 洋子

第2セッション 14:20 - 15:10

司 会 名古屋大学 山下 淳子

3. "Maybe," "Perhaps," "Probably" は日本人英語学習者の書き言葉でどのように使用されているか
学習者コーパスを利用したアプローチ
昭和女子大学 小林多佳子
4. 専門英語教育におけるミニコーポラの有効性
武庫川女子大学 野口ジュディー

休 憩 15:10 - 15:20

シンポジウム (1) 15:20 - 18:00

日本における英語コーパス言語学の現状と展望

コーパスと語彙	司 会	京都外国語大学	赤野 一郎
コーパスと文法	講 師	京都外国語大学	赤野 一郎
コーパスとテキスト	講 師	椋山女学園大学	深谷 輝彦
コーパスと英語史	講 師	大阪大学	田畑 智司
コーパスと英語辞典	講 師	山口大学	西村 秀夫
コーパスと英語教育	講 師	徳島大学	井上 永幸
	講 師	東海大学	朝尾幸次郎

懇親会 18:10 - 20:00 会費 4,000 円 司 会 名古屋大学 滝沢 直宏
(懇親会の申し込みは 10 月 3 日午後 5 時までに E-mail あるいは葉書で事務局へ。会費は当日お支払い下さい。)

10 月 6 日(日)

ワークショップ (2) 9:00 - 10:30

《正規表現によるテキスト検索 (応用編)》 講 師 名古屋大学 大名 力
定員 先着 40 名 (予定) 参加費 会員無料・非会員は当日会費 1,000 円 (申し込みは電子メール・郵便で事務局まで) なお、ワークショップ(2)の受付は 1 日目のワークショップ(1)の受付とは別に行っています。どちらを受講するかを明記してください。両方受講することも可能です。

休 憩 10:30 - 10:50

シンポジウム (2) 10:50 - 12:50

コーパスを利用した英語教育と英語・英文学研究指導 実践報告と今後の可能性

コーパスを英語教育に生かす	司 会	関西学院大学	八木 克正
コーパスを現代英語研究の共時的研究指導に生かす	講 師	明海大学	投野由紀夫
コーパスを英語の通時的研究指導に生かす	講 師	帝塚山大学	梅咲 敦子
コーパスを英文学研究指導に生かす	講 師	中部大学	大門 正幸
	講 師	広島国際大学	石川慎一郎

休 憩 12:40 - 13:30 (昼食はご持参下さい)

特別講演 13:30 - 15:30

Corpus linguistics: past, present, future	司 会	徳島大学	中村 純作
	講 師	University of Oslo	Stig Johansson

閉会の辞 名古屋大学 杉浦 正利

英語コーパス学会

(Japan Association for English Corpus Studies)

会長 今井光規 事務局 770-8502 徳島市南常三島町 1-1 徳島大学総合科学部 中村純作研究室
TEL: 088-656-7129 E-mail: jun@ias.tokushima-u.ac.jp 郵便振替口座 00940-5-250586
URL http://muse.doshisha.ac.jp/JAECS/index.html

大会当日、入会受付もいたしますので、お誘い合わせの上ご参加下さい(年会費 一般 5,000 円 学生 4,000 円)。
また「当日会員」としての参加も受け付けております(1,000 円)。

英語コーパス学会創立 10 周年記念第 20 回大会レジュメ

ワークショップ《正規表現によるテキスト検索》

(講師 大名 力)

電子化テキストを利用した言語研究において、テキスト処理の技術は不可欠なものであり、テキスト処理において正規表現の果たす役割は大きいですが、実際には、「任意の一字」「0/1 回以上の繰り返し」程度のもしか利用されていないことが多い。本ワークショップでは、正規表現の初歩から始め、「worth の後に、0 語から 3 語挟んで、ing で終わる語(ただし、thing, something, anything, everything, nothing は除く)が続くもの」のような、複雑な条件の指定の仕方まで見ていく。

初日のワークショップでは、これまでに正規表現を使ったことがない人を対象に、基本的な正規表現の使い方について見ていく。2 日目は、初日の内容を前提に、少し複雑な条件の指定方法について見ていく。どちらか一方のみの参加も可能。どちらも特別な知識 (UNIX や Perl の使い方など) は必要としない。内容は次の URL で確認することができる。

<http://infosys.gsid.nagoya-u.ac.jp/~ohna/re/>

研究発表 第 1 室

アメリカ英語における国名の使用頻度に基づく「認知的世界地図」構築の試み

(松野 和子)

この研究は、コーパスを用いて、事物への印象など人の持つ認知的枠組みを明らかにすることを目的としている。本発表では、そのためのパイロットスタディーとして、Brown コーパスと Frown コーパスをデータとし、国名の頻度、国名と共起した名詞、国名が現れるジャンルを分析した結果を報告する。共起した名詞やジャンルを分析することによって、それぞれの国を座標軸上に表し、国の位置関係を明らかにするとともに、頻度の多さを各国への関心度の高さとみなし、それを面積でグラフ上に表す。こうしてできた図を本研究では「認知的世界地図」と呼ぶが、Brown コーパスと Frown コーパスを比較することにより、1961 年から 1991 年の 30 年間にそのような「認知的世界地図」がどのように変化したかを明らかにする。

分析は次のような手順に従って行った。

- 1) 世界約 180 カ国の国名の頻度を調べる。頻度の低いものは分析対象から外した。
- 2) 国名を含む文中に出現する各名詞の頻度を調べる。国名自体の頻度による影響を考え共起する名詞の相対頻度を出す。
- 3) コーパスを構成する 15 ジャンルごとに国名の頻度を調べる。ジャンルごとの総単語数が異なるため 5 万語あたりの相対頻度を出す。
- 4) 共起した名詞・ジャンルという 2 種類の多元的データを用いて、それぞれコレスポネンス分析を行い、国の位置を決めるカテゴリースコアを算出する。
- 5) カテゴリースコアをもとに、共起した名詞によって国の位置関係を表したものと、ジャンルによって国の位置関係を表したものという 2 種類の「認知的世界地図」を作成する。その際、国の延べ頻度を換算して、地図に面積を付け加え「認知的世界地図」を完成させる。
- 6) 「認知的世界地図」と見比べながら、各国への印象を具体的に考察するため、共起した名詞とジャンルそれぞれのサンプルスコアを座標軸上に示す。

上記の手順を、Brown コーパスと Frown コーパスに関して行い、結果的に、1961 年と 1991 年においてそれぞれ、共起した名詞によって印象を明らかにした「認知的世界地図」とジャンルによって印象を明らかにした「認知的世界地図」ができあがる。本研究では、この 4 枚の「認知的世界地図」を比較分析することにより、アメリカ英語に現れた各国の認知的位置関係と関心の強さの移り変わりを考察する。

アカデミックスピーカーの発話語数のジェンダー差

Corpus of Spoken Professional American-English の分析から

(家口 美智子・家入 葉子・岡部 浩子)

英語を母語とする話者のジェンダー差については 1970 年代以来いろいろなアプローチによる研究分析が盛んに行なわれるようになった。先行研究は往々にして会話参加者が少なく観察時間が限られている小規模な

会話場面の分析にとどまっている傾向があるため、我々は Barlow (2000) による Corpus of Spoken Professional American-English (以下 CSPAE) (総語数 200 万語) を利用し、ジェンダー差を大規模な会話場面から検討するためのプロジェクトを開始した。本研究はその第一歩であり、「男女のどちらがよくしゃべるか」という問題について発話語数という点からのジェンダー差の分析を行った。

Tannen (1990) が「私的な場では女性が公的な場では男性がよくしゃべる」という理論を展開している。公的な場の会話を調査したどの計量的先行研究もほぼ男性優位の結論を出している。我々は CSPAE の大学教員の会議 (90 万語) を利用して、延べ 381 名 (男性: 209 名; 女性: 172 名) に上るアメリカ英語のアカデミックな話者の公の場での発話語数を分析した。コーパスを使用しているためこのトピックの先行研究よりも大規模の調査であり、ある程度の客観的な結果を期待できると言えよう。

このコーパスの大学教員の会議は Faculty 会議, Mathematics 委員会, Reading 委員会の 3 つのセッティングからなる。1 回限りの会議ではなく司会者, 出席者とも若干の入れ替えはあるものの似たようなメンバーが集う連続した会議のうち数回が録音されている。Faculty 会議, Mathematics 委員会ではどの会議でも男性話者が, Reading 委員会ではどの会議でも女性話者がグループとしても一人あたりの発話としても発話語数が一貫して多い (司会者の発話は分析からはずしている)。これは従来説明されてきた「私的な場では女性が公的な場では男性がよくしゃべる」(Tannen, 1990) や「アカデミックな女性は男性と差はない」(Lakoff, 1975) という見解のどちらも支持しない。さらに会議参加者が多い性のグループの方が一人あたりの発話語数も多い傾向にあることが明らかになった。また、発話語数において各セッティングのどの会議においても一貫した男性優位, または女性優位の傾向が見られることは会議における会話スタイルを考察する上で、語彙や文体同様重要な点を示唆していると考えられる。

以上, CSPAE を使った発話語数のジェンダー差の分析は, 3 つのセッティングを対象とし客観的な計量的手法に基づいているためアメリカ英語知的話者の会話スタイルをある程度客観的に提示したと言えよう。

文の直接構成素レベルでの解析済コーパスの比較

(田中 泉)

本研究は, 2 つの異なったジャンルの解析済コーパス (parsed corpora) を文の直接構成素 (Immediate Constituent) レベルで比較した際の観察報告である。ここで言う解析済コーパスは, Skeleton Parsing Scheme という句構造 (phrase structure) 分析のガイドラインに沿った, 英文コーパスの句構造解釈の記録である。そこでは, 例えば平叙文 (S) の直接構成素は, 多くの場合, 主語に相当する名詞句 (NP) と述語に相当する動詞句 (VP) に分けられる。(S NP, VP) このように文の直接構成素を見れば, その文の大まかな構造, タイプを知ることができる。もしコーパスの特徴が, 文の構造のレベルにおいて認められるとすれば, それは, 何らかの形で句構造解釈の記録である解析済コーパスに反映されることが予想される。使用した HANSARD コーパスはカナダ議会の議事録から AP コーパスはアメリカ Newswire Reports から採取したものである。

直接構成素のパターンを比較観察した結果, 2 つのコーパス間で頻度の差が比較的大きいパターンの中に, それぞれのコーパスの特徴が反映していると考えられる。幾つかのパターンを認めることができた。(以下, 例文の後の括弧内に直接構成素のパターンを示す。)

カナダ議会の議事録である HANSARD コーパスには, 次のような議会特有の発言, 言い回しが多く見られる。

- “Order, please.” ([N N], RR)
- “Bill C-123, to amend the Criminal Code ...” ([N N], [Ti Ti] ...)
- “Mr. Speaker” 等の呼びかけを含む文 (e.g. [N N], [N N] [V V].)
- “in view of” をつけた表現 (e.g. [P P], VM [N N] [V V]?)

AP コーパスはニュース記事であるから, 直接・間接話法の文, reporting clause の頻度, 特に, 文末の reporting clause の頻度が高い。例としては次のような文が挙げられる。

The weather can change in a hurry in the mountains, he said. ([N N] [V V], [Si Si].)

また, 次の例は北米のニュース記事によく見られる構文である。

Among the organizers was consumer advocate Ralph Nader. ([P P][V V].)

このように、解析済コーパスから文の直接構成素を抽出・比較することは、そのコーパスのジャンルや文の構造上の特徴を知る上で有益であると考えられる。

文体差に関わる語彙の特定 Bank of English の「大衆紙語」と「高級紙語」の形容詞 (高見 敏子)

文体の構成要素の一つとして、語彙が果たす役割を無視することはできない。従来の比較文体研究では、(1)どの語を(2)いくつまで、比較対象として選択するかという2点について、研究者の主観で判断されてきた。しかし、この方法では明確な基準がないまま部分的選択がされることになり、本来は同程度あるいはそれ以上に注目すべき語が見過ごされてしまう可能性は否定できない。文体の差に関わる語彙を、複数のコーパスの語彙頻度データを使って客観的かつ網羅的に導き出すことはできないだろうか。本研究は第一義的には方法論としての、corpus-drivenな語彙分析の試みである。

対象としたのは、2000年9月のBank of Englishに含まれていた、2つの大衆紙と3つの高級紙からなる新聞サブコーパスである。イギリスの大衆紙と高級紙は文体の比較対象としてしばしば取り上げられるテキストであり、使用語彙の差の重要性は関連する先行研究でも必ず指摘されてきたにもかかわらず、研究の中心的対象にはされてこなかった。本研究の第二の目的は、この二種類の新聞における語彙の差を(その有無も含めて)、データによって提示することである。

使用語彙の差を特定する統計量として、POS タグ付き word form の頻度について Dunning (1993)の log-likelihood statistic を計算し、使用頻度の差の有無の判定基準として 7.88 (カイ二乗分布における、自由度 1, 有意水準 0.5%の値) を採用した。この統計量は、特に中・低頻度の語彙の distinctiveness について、カイ 2 乗統計量や MI よりも妥当な結果が得られるとされている (Daille, 1995)。本研究で得られた結果から、コーパスサイズや語彙の頻度の高低といった要因が log-likelihood statistic に与える影響につい

ても検討する。

研究を進める上で、各段階で可能な限り数値的な根拠に基づいた判断をするように努めた。出発点として、5つの新聞を始めから大衆紙と高級紙という対立する2つのグループに分けて考えるのではなく、全てが独立した新聞と捉え、リーグ戦方式に計10組の全ての組み合わせについて、上記の計算を行った。この「総当り比較法」の実践上の詳細については発表で説明するが、この方法で「有意差パターン」「差の方向パターン」を調べたところ、まず使用語彙の差について大衆紙と高級紙との区分は妥当という結論が得られ、次に大衆紙・高級紙にそれぞれ相対的に多い語が特定された。

以上の手続きで得られた語を「大衆紙語」「高級紙語」と呼ぶことにする。名詞・動詞・副詞など、その品詞はさまざまであるが、本発表では形容詞を取り上げる。結果からは、先行研究で触れられた特徴以外の形態論的・意味論的特徴も浮かび上がった。特定された語が現れるコロケーションやコンテキストの例を検討し、それぞれの語彙使用の差が生じる背景を、内容・文体の両要因から分析する。

研究発表 第2室

国際学習者コーパス ICLE/LINDSEI 最新の動向

(金子 朝子)

本発表では、スウェーデンのイエテボリで行われた ICAME2002 期間中の ICLE/LINDSEI 各国代表者会議でのトピックを中心に、様々な言語背景を持つ英語学習者の中間言語比較を可能にしてくれる ICLE/LINDSEI の動向を紹介する。

ICLE (International Corpus of Learner English) は、ロンドン大学の Greenbaum によって始められた世界 18 地域の英語変種コーパス構築のプロジェクトである ICE (International Corpus of English) の傘下で、1990年にスタートした。ベルギーのルーバン大学の Granger を中心に世界 18 カ国以上のチームによって運営されている。異なる母語を持つ、英語を外国語として用いる大学の上級生 (3・4年生) の argumentative essay を各国 20 万語以上集めたものである。現在は ICLE プロジェクトのメンバー以外には公開されていないが、本年秋には、ベルギー、スウェーデン、ロシア、ブルガリアなど 11 カ国のデータに被験者のプロファイルの検索や簡単な分析のため

のツールをつけた CD-ROM が発売される。残念ながら日本を含む、アジア諸国のサブ・コーパスはこれには含まれていない。しかし、近い将来、日本や他のアジア諸国、更にヨーロッパの国々からのデータを加えた CD-ROM 第 2 版が発売される予定である。また、英語母語話者の同様のデータを集めた LOCNESS コーパスは ICLE 日本代表を通して一定の条件のもとで利用可能である。すでに、各国のサブ・コーパスを利用して、また、LOCNESS との比較研究など盛んに研究が進められており、ルーバン大学の英語コーパス言語学センターのサイト(www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/cecl.html)にはこれまで研究のリストも掲載されている。CD-ROM の発売に続いて、メンバー国では各国のデータにエラータグをつけたものを作成するプロジェクトを開始する予定である。

1995 年には ICLE と平行して学習者の話しことばを集めたコーパスの収集が始められた。LINDSEI (Louvain International Database of Spoken English Interlanguage) はフランス語を母語とする大学上級生(3・4年)英語学習者のデータに続いて、話しことばのコーパスに興味を持つ ICLE メンバー国が、決められたデータ収集方法でサブ・コーパスを作成中である。現在、EFL データとしてフランス、イタリア、日本、スペイン、スウェーデンが当初目標としていた 50 サンプル以上の収集をすでに終えている。コロンビア大学での ESL 日本人データや、母語話者のデータも整えつつある。LINDSEI はまだ収集途中で、プロジェクトメンバーのみへの公開となっている。日本人サブ・コーパスは、更にサンプルを加え、来年の春に Web 上で公開する計画がある。

CD-ROM の公開によって、Granger が目的としていた、異なる L1 を持つ学習者のエラーを比較することが可能となり、universal error と transfer error を特定することも不可能ではない。ICLE/LINDSEI が様々な中間言語研究により良いデータを供することができることを期待している。

英語学習者コーパス分析法の開発から電子掲示板の改良へ

(鈴木 千鶴子・吉原 将太・渡辺 洋子)

発表者らは、日本語母語話者の英語コミュニケーション力養成を目的としたインターネットコミュニケーション実践を促進する e-learning 環境の開発・構築に着手

し、4年目を迎えた。その間に、学内サーバーに設置したインターアクティブ・ウェブサイト上の電子掲示板に書き込まれた投稿送信データを学習者コーパスとして、約 300 名分延べ約 3000 件を集積した。

この学習者コーパスに基づき、本ネットワーク上の e-learning 環境を評価する限り、従来の教室授業環境で不足しがちな双方向コミュニケーションを容易にし、併せて教室外、授業時間外での学習機会を提供し、結果として自立的・自律的学習を支援するものであること、特に、流暢さ (Fluency) を涵養するものであることが確認された。しかしながら、語学上の誤りも数多く見られ、教師やネイティブ・スピーカー側からの矯正を意図した介入にも拘らず、同じ誤りが繰り返され、せっかくのコミュニケーションの機会が、誤りの再強化の機会と化すことも懸念される結果を観察した。

そこで、流暢さならびに自立的学習の養成を損なうことなく、正確さ (Accuracy) の学習を促進する支援ツールを如何に付加するかが、本 e-learning 環境の進化の課題となった。その設計にあたり、理論的根拠を提供するものとして学習者コーパスの分析、とくに誤りの分析の方法が検討された。一案として、学習者コーパスにエラータグを付け WordSmith 等によるエラー分析が試行されたが、作業の手間と成果を勘案し今回は見送られた。代わって、オリジナルな「語句検索システム」を、教師側に限らず学習者自身も利用可能なウェブアプリケーションとして開発した。

投稿前の段階でメッセージ確認ページにエラーの可能性のある個所を警告表示するシステム(エラーチェックシステム)の導入が並行して試行されているが、オリジナルに開発した本「語句検索システム」を利用した Accuracy 習得効果に対する期待の方が大きい。その理由は、ミクロな局所的エラーチェックに止まらず、当該学習者コーパスが言語学習素材としてディスコースレベルでの発見学習を促進するものであるためと考えられる。発表では、本システムの概要と使用成果の一部を紹介し、広く意見を仰ぎたい。

“Maybe,” “Perhaps,” “Probably” は日本人英語学習者の書き言葉でどのように使用されているか 学習者コーパスを利用したアプローチ

(小林 多佳子)

本発表の目的は、学習者コーパスを利用して事実の可

能性の段階を表す副詞 “probably,” “maybe,” “perhaps” がどのように日本人英語学習者の書き言葉で用いられているかを調べることである。

Leech (1975: 218-219) によれば、私たちは真実や虚偽を黒白ではっきりと区別して考える代わりに、「ありそうな程度の段階 (SCALE OF LIKELIHOOD)」という観点で考えることができ、これらの概念は、さまざまな方法で表現されるという。最も重要なのは法助動詞 can, may, must などによる方法であり、さらに “It is possible that you are right” のように先行の it や that 節を伴う文による方法、さらに “perhaps,” “probably” などの副詞類を用いる方法である。

本研究では文全体の内容に対する、書き手自身のコメントを表す上記の三つの副詞が日本人英語学習者にどのように使用されているかを分析し、以下の3点を中心に考察する予定である。

- 1) 上記の三つの副詞の使用頻度は日本人学習者と母語話者の間で違いがみられるか。
- 2) 日本人英語学習者のこれらの副詞の使用には文の構造上、顕著な特徴が存在するのであろうか。
- 3) これらの副詞の使用には日本語からの転移は見られるのであろうか。

データとして使用するのは 146,692 語からなる 289 人の日本人上級英語学習者が書いた、各 500 語以上の英語の argumentative essay である。比較研究のために 176 人のアメリカ人大学生の書いた 144,853 語からなるサンプル・コーパス (LOCNESS) も母語話者のデータとして使用した。この研究では WordSmith 分析ツールの Concord を使用してこの三つの副詞がどのような文脈で用いられているかを検索し、日本人英語学習者と母語話者のデータを比較する。

専門英語教育におけるミニコーポラの有効性

(野口ジュディー)

EFL (English as a foreign language) におけるコーパスの利用の有効性は言うまでもないが、実際に大学の共通教育あるいは一般教育で利用するには与えられる英文が専門的過ぎる、あるいは学生には難しすぎるといった問題点がある。そこで、ここでは言語や文系以外の理工学専攻の大学院生の専門英語教育を例として取り上げ、ESP (English for specific purposes) におけるミニコーポ

ラの利用を提案したい。

専門英語は一般に使用されているものと違うので、一般用の辞書や参考書が役に立たないことが多い。専門用語の使用にあたって、たとえば、血液学の分野で cancer と tumor と neoplasia をどのように使い分ければ良いのかについては、一般辞書からは答えは出てこない。また、一般辞書には載っているが、専門分野の文書を作成するにあたって単語の選択に困ることが多い。たとえば、indicate/show/suggest/reveal をどのように使い分けるかについては、一般の辞書やコーパスからでは適切な選択のヒントは得られにくい。

このような問題の一つの解決策としては、教師が学生の専門分野に合わせてコーパスを準備し、分析して、学習問題を作って教室で教えるといった方法が考えられる。しかし、このようにすることには二つの問題点がある。一つは ESP 教育の目標である学習者が実際に自分の専門分野で活動するときに役に立つ言語力を身につけさせることが、この方法では必ずしも充分達成できないことである。もう一方の問題点は、教室には同じ専攻の学生がいても、異なる研究に携わっているため専門用語が違うことである。たとえば、同じ知能工学でもコンピュータシミュレーションを中心に研究している学生もいれば、実際に模型を作って実験をしている学生もいる。これらの条件を考えると、ESP の教室では学生自身にコーパスを準備させ、その使い方を教えることが、将来一番役に立つ教授法ではないかと思われる。

ミニコーパスを使った授業の一例として大学院生を対象とした論文の書き方のクラスが挙げられる。学生には、まず ESP 研究で明らかになった論文の特徴を説明したうえで、自分の研究のために読んでいる論文からこれらの特徴を持つ文章を探させ、電子テキストで提出してもらおう。全ての学生のテキストをプールして Exercise を作り、それに基づいたディスカッションを教室で行うが、ここでは discourse signal や collocation にも注目させている。このように学習したテキストの特徴を利用して、学生は最後に自分の研究を説明する文章を完成することになる。ESP 教育の狙いは完成品だけではなく、そこにたどりつくプロセスも重視することにあるので、検索作業の結果や教室でのディスカッションへの参加の度合いが評価の中心となるが、学生の作成した文章も評価の際には考慮している。

このアプローチを使い、理工系の大学院で化学工学、生物学、機械工学、知能・機能創生工学などの学習者に

専門英語を教えて、学生からも高い評価を得ている。また、学生自身が書いた英文を簡単なラーナーコーパスに編纂し、それを利用することにより、英語学習者の悩みである冠詞の使い方や、専門英語の基礎的文法的な特徴の学習にも役立たせている。本発表では、具体的なミニコーパスの作成法や、その利用法を紹介しつつ、ESPにおけるコーパス利用の有効性について論じたい。

シンポジウム (1)

《日本における英語コーパス言語学の現状と展望》

(司会 赤野 一郎)

1990年5月、近代英語協会第7回大会において齊藤俊雄氏の司会によるシンポジウム「近代英語とコンピューター」が開かれ、現在の意味での「コーパス」という言葉が初めて使われた。その2年後の1992年11月、日本英語学会10回大会において吉村由佳氏が立案したワークショップ「Corpus Linguistics - その理論と応用 -」が、コーパス言語学を全面に出したわが国で初めての試みであった。その内容は次の通りである。

司会 吉村由佳

講師 藤本和子 「コーパス構築とその問題点 - コンピュータコーパスへの招待」

杉浦正利 「接続語句 however の生起位置分析 - コーパス利用の実際」

清水 誠 「コーパスを用いた束縛理論の検証」

吉村由佳 「コーパスに見る reason why - その文脈と使用条件」

この2つの活動が合流し、翌年、大阪大学(大学院言語文化研究科)で、齊藤俊雄氏を会長に当学会の前身、「英語コーパス研究会」が発足し、名称の変更を経て今年、10年を迎えたわけである。

このシンポジウムでは、今日に至るわが国の英語コーパス言語学の主要な成果を概観することによりコーパスが英語研究にいかにか寄与したかを検証し、あわせて今後の見通しを探ることを目的とする。各講師には、限られた時間であるが、語彙、構文、テキスト、英語史、辞書、英語教育のそれぞれの領域との関わりで、主要な研究を概観・整理し、今後の見通しを語っていただく。取り扱う領域によっては、検討期間をさらに過去へ広げる必要があるだろうし、国内のみならず海外に目を向けてこそ、その領域全体の成果が見通せる場合もあることを

お断りしておく。

コーパスと語彙

(赤野 一郎)

コーパスを検索するためには「コンコーダンサー」と呼ばれる、検索語を行の中央に配置し、その左右に文脈を添えた1行単位の検索結果、KWIC コンコーダンスを生成するプログラムが必要である。コーパスを用いた語彙研究は、このKWIC コンコーダンスの精査と、語の出現頻度を中心とする統計処理につきると言っても過言ではない。

検索結果を解釈するためには分析者に鋭い直観と洞察力が要求されるが、KWIC コンコーダンスの最大のメリットは、それまで気づかなかった語の様々な振る舞いのパターンを我々の目に見える形で提示してくれたことであり、その点で語のコロケーション研究が大幅に進歩し、辞書編纂方法に直接的影響を与えた。

さらに、一見単純なこの分析方法は、新たな言語観形成に多大な影響を与えた。Sinclair (1991) *Corpus, Concordance, Collocation* において、このコンコーダンスラインの精査による言語使用の分析を通じて、言語使用者は文法規則によって文を作り出しているという考え方では説明がつかないほど、実際の語の結合は限られていることを示した。つまり言語は可能な単語の組み合わせがすべて実現されているというより、大半はあらかじめ定型化された (pre-constructed)、予測可能なことばのかたまりからなっているという説が展開された。この説は外国語教育でも支持され、Lewis 等の Lexical Approach の理論的裏付けとなっている。

一方、統計処理による頻度分析はコンピュータのもっとも得意とするところであり、語彙研究以外のここで採り上げるすべての領域において、採用されている手法であるが、留意すべきことは数量的分析のみに終わってはならない、数値の意義付け、すなわちいかに深い質的分析がなされたかである。

本発表では、コンコーダンスと統計処理を活用した語の分析に新たな地平を切り開いたと評価できる国内外の研究を、主として『英語コーパス研究』、*ICAME Journal*, *International Journal of Corpus Linguistics*, *English Corpus Linguistics in Japan* から採り上げ、コーパスに基づく語彙研究の動向を概観し、その有効性を提示したい。

コーパスと文法

(深谷 輝彦)

この発表では、過去 10 年間に日本で行われたコーパスに基づく英文法研究を数量的に概観する。調査対象として『英語コーパス研究』に掲載された論文を中心に利用するが、必要に応じてその他の学会誌や著書のコーパス英文法研究も含める。また比較のために、同時期の *ICAME Journal* も調べる。

第一に、1960 年から 1993 年のコーパスによる英語研究を分類した Kennedy (1998: 89-90) の枠組みを用いながら、どの文法項目がよく研究され、どの文法項目が手つかずになっているか、数量的に提示する。その上で、議論したいのは研究の濃淡が生じる原因である。すなわち、

- 1) 研究者が関心をもつ英語構文を調べるのに適当なコーパスがあるか否か、
- 2) コーパスから当該の構文を検索することが可能かどうか、
- 3) その構文について文法的一般化を行なうのに十分な量のデータが得られるか、
- 4) データに統計処理等を施して構文の特徴を明示的に示すことができるかどうか、

という論点をもとにこれまでの研究をふりかえる。

第二に、何のために文法研究にコーパスを利用するのか、コーパスを利用した結果、どのような創造的な研究が可能なのかという基準でこれまでの研究を整理する。園田 (2002: 271) は、「コーパスに対して行う言語分析は煎じつめると、頻度 (frequency) の調査と分布 (distribution) の調査の 2 つになります。」と指摘する。その具体的内容として、類似の機能を果たす構文を比較する、語彙と構文の連続性を解明する、言語使用域により変異研究 (例、話し言葉と書き言葉の比較) を行う等の調査をすると、コーパスはその威力を発揮する。また、文法理論構築の際の事実観察、あるいは理論の追試験としてコーパスがはたす役割も大きい。ここでは、コーパスを何のために用いているのか、という論点から文法研究を分類する。

最後に、今後のコーパス利用英文法研究の展望として、従来の語彙に基づく構文検索に加えて、構文解析を施したタグ付きコーパス検索の普及がコーパス利用文法研究の幅を広げるために必要であることを論じる。(cf.

Meyer (2002)) 構文研究をする時、プレーンテキストをコンコーダンス・ソフトに読み込んでキーワードで構文を検索する方法が従来多く利用されたきた。しかしこの研究方法では、膨大なバグを含んだ検索結果を生み出す構文や語彙的指定が難しい構文は扱いにくくなる。そこで Meyer (2002) が実践しているように、文法情報付きコーパスを活用することで新たな文法研究領域を開拓できることを示し、まとめとしたい。

コーパスとテキスト

(田畑 智司)

テキストにおける語彙の生起頻度等の統計値をもとに文体や著者推定、テキスト・タイポロジー、使用域による言語変異の問題を扱った研究の歴史は比較的長く、電子コーパスの出現以前にまでさかのぼることが出来る。Ellegård (1962) による 'Junius Letters' の著者推定や Milic (1967) による Swift の文体研究などは先駆的業績の代表例である。テキスト中の語彙や構文等、言語項目の生起頻度を集計し、他のテキストとの「比較」を行うことで、文体的特徴や言語特徴を探り出すことを基本とするディシプリンにおいては、電子テキストやコーパス、そしてコンコーダンス等の言語処理プログラムの出現はまさに発展の起爆剤となった感がある。実際、そうした環境が整い始めた 1980 年代半ばから計量研究の手法も一段と進歩し、Burrows (1987)、Biber (1988) などをはじめ、電子テキスト・コーパスを活用し、高度な統計学的手法を適用するアプローチが本格化してきた。

国内の業績に目を転じると、いち早く多変量解析を取り入れた Nakamura (1991, 1992) による LOB/Brown コーパスに見られるジャンル・スタイルの研究をはじめとして、1990 年代以降、数量化 III 類によって Bank of English、BNC など主要コーパスを次々と分析対象とした同氏の研究は海外でも広く知られているところである。同じく数量化 III 類を用いた高橋 (1995) および Takahashi (1997, 1999) は LOB コーパスや BNC を構成するテキスト・ジャンルの類型化を試みたものであり、また発表者は主成分分析やコレスポネンス分析によって高頻度語や文法範疇の生起頻度分布を解析して Dickens の作品の会話文体の特徴や文体の経年変化を取り扱っている。さらに、クラスタ分析を適用して音声言語および文字言語のレジスターにおける統語現象の生起頻度を解析した一連の Umesaki 論文 (1994, 1997, 2000a/b, 2002)。構文解析標

識が付与された Penn Parsed Helsinki Corpus を対象とした Tsukamoto (2002) もクラス分析を利用したテキスト分類研究の成果である。

他方 統計的手法によらないテキストおよび文体へのアプローチについても取り上げるべき業績が出てきている。ルイス・キャロルによる一連のアリス小説の文体を考察した稲木・沖田両氏による共著書 (2002)、副詞のコロケーションを通して Dickens の文体の諸相を解き明かそうと試みる Hori (1999, 2002) など、電子テキストを効果的に活用した事例であろう。

本発表ではこのように、テキストおよび文体の問題に関して、これまで国内外でなされてきた研究の潮流や動向を俯瞰し、分類・整理しながら、コーパスに基づくテキストへのアプローチの可能性や限界、今後の展望について時間の許す範囲で論じたいと思う。

コーパスと英語史

(西村 秀夫)

英語史研究の領域では、1980 年代に入って、研究の道具としてコンピュータがどの程度有用かということに対する関心が高まった。その一つの大きな流れを、コンピュータによるコンコーダンス編纂に見ることができる。英語史研究では、原典資料を読みながら必要な用例を見つけ出すという作業が欠かせない。これを効率的に行なうための手段として、英語史研究者がコンピュータに注目したのは当然のことであった。この時期の代表的な業績としては、大型計算機を利用した Osaka University Concordances to Middle English Metrical Romances シリーズの刊行や、パーソナルコンピュータで容易に KWIC コンコーダンスや語彙リストを作成することを可能にしたソフトウェア IDEA-III の開発などを挙げることができる。また日本英語学会第 6 回大会 (1988 年) では、「コンピュータ・コンコーダンスで何ができるか」と題するシンポジウムが行なわれた。

1990 年 5 月に開催された近代英語協会第 7 回大会で「近代英語研究とコンピューター」と題するシンポジウムが行なわれた。その内容は次のとおりである。

司会 齊藤俊雄

講師 中村純作「Brown Corpus 等のコーパス利用による英語分析」

今井光規「パソコンを使って用例カードを作る」

久屋孝夫「パソコンを利用した初期近代英語の基礎的研究」

このシンポジウムは次の 2 つの点で重要な役割を果たしたと考えられる。

- 1) このシンポジウムが契機となって、古英語期から近代英語期までをカバーする論文集『英語英文学研究とコンピュータ』(齊藤俊雄編 英潮社 1992 年) が刊行され、コンピュータを英語英文学研究に利用する研究者の裾野が広がったこと
- 2) 司会の齊藤教授から、当時まだ編纂中であった Helsinki Corpus に関して詳しい紹介があり、英語史研究者がコンピュータコーパスについて本格的に知る初めての機会となったこと

Helsinki Corpus は 1991 年に公開された。1993 年には英語コーパス研究会(現英語コーパス学会)が発足した。学会組織ができたおかげで通時的英語コーパスに関心を持つ研究者のネットワークが形作られ、Helsinki Corpus その他のコーパスを利用した言語研究が大きく進展した。

今回のシンポジウムで私に課せられた役割は、コンピュータコーパスと出会ったことで、英語史研究がどのように変貌し、進化してきたかを、ここ 10 年間の研究の動向を概観しながら示すことである。

コーパスと英語辞典

(井上 永幸)

600 万語の書き言葉と 130 万語の話し言葉を Birmingham Collection of English Text (後の Bank of English) から抽出したものを基に編集された Collins COBUILD English Language Dictionary (HarperCollins; COBUILD) が 1987 年に発刊されて以来 15 年になる。その後、1995 年には Cambridge International Dictionary of English (CUP)、Collins COBUILD English Dictionary (Second Edition, HarperCollins)、Harrap's Essential English Dictionary (Chambers Harrap Publishers)、Longman Dictionary of Contemporary English (Third Edition, Longman; LDOCE3)、Oxford Advanced Learner's Dictionary of Current English (Fifth Edition, OUP; OALD5) など(以上、アルファベット順)、改訂版と新刊を含めて各種出版社からコーパスに基づくことを標榜する ESL/EFL 辞典が次々と刊行さ

れた。また、その後も、OALD6 (2000, Sixth Edition; OUP), LDOCE3² (2001, Third Edition with New Words Supplement; Pearson Education), *Collins Cobuild English Dictionary for Advanced Learners* (2001, Third Edition, Harper-Collins) などの改訂版が出版されている。一方、*Cambridge Dictionary of American English* (1999, CUP) が先鞭をつけて以来、アメリカ英語に主眼を置いた ESL/EFL 辞典の出版も盛んになってきた。*Longman Advanced American Dictionary* (2001, Pearson Education Limited) や *Macmillan English Dictionary* (2002, Oxford: Macmillan Education; MED) がそれである。さらに、最近刊行された本格的なコロケーション辞典、*Oxford Collocations Dictionary for Students of English* (2002, OUP) も注目されるところである。

コーパスに基づく辞典の編集方針にも大きく二つの流れが出てきた。一つは COBUILD や LDOCE3 に見られるような、コーパスにより明らかになった典型的な「形」に注目して語義を記述してゆく言わば「形に基づく編集法」である。もう一つは OALD6 や MED に見られるような、形にも注目するものの語義分類自体は従来の語義のまとまりを中心に記述してゆく「意味と形の融合による編集法」である。後者は、COBUILD 発刊以来、新潮流となっていた前者の編集方法に真っ向から挑む形で提案された編集方法といえる。

日本においても、大辞典では「編集にコンピュータコーパスを利用した」とし、「語や表現の頻度、用例、語法記述の参考資料とした」ことを明記したり、大辞典としては初の試みである頻度表示にコーパスを利用したことを述べる『ジーニアス英和大辞典』(2001, 大修館書店)や「コーパス言語学の資料を活用して、重要語を中心として語義などを原則として頻度順に配列し」とする『新英和大辞典』(2002, 第6版, 研究社)、中辞典では「見出し語はもちろん用例や日英比較など」の点で「電子コーパスやウェブサイトなども参照」したとする『ルミナス英和辞典』(2001, 研究社)、頻度表示にコーパスを利用したとする『ジーニアス英和辞典』(2001, 第3版, 大修館書店)など、コーパスの利用を謳う辞書も現れ始めてきている。本発表においては、コーパスに基づく英米の EFL/ESL 辞典の歴史を振り返りながら最近の動向を概観し、適宜国内の英和辞典の今後の流れを占ってみたい。

コーパスと英語教育

(朝尾 幸次郎)

この10年間におけるコーパスと英語教育における大きな流れはふたつある。ひとつは英語学習へのコンコーダンスの利用、もうひとつは学習者コーパスを利用した研究である。

1) コンコーダンスの利用

言語研究に利用されてきたコンコーダンスという手法を語学教育に利用しようとする考えは1980年代後半に現れ、90年代に広く認知されるようになった。教師がことばの意味や使い方を知識として学習者に教えるのではなく、学習者みずからがデータをもとに自立的に学ぶ data-driven learning という考え方にそったものである。

わが国では杉浦正利氏(名古屋大学)が WebGrep for NESS という名前で、英語学習用英日対訳例文集を公開されている。この特徴はアメリカ人、カナダ人インフォーマントによるオリジナルな英語例文集をコーパスとしたもので、それぞれに日本語訳がつけられている。一般のコーパスで検索した結果は学習者にとって用例がむずかしすぎたり、検索例が多すぎるという不便がある。WebGrep for NESS は学習者向けの用例が検索できるだけでなく、パラレルコーパス的に英日、日英のどちらからでも検索できる点で、外国語学習へのコンコーダンス利用の新しい道を開くものである。

2) 学習者コーパス

1990年代、言語研究を目的とした大規模コーパスが次々と開発されるなかで、学習者コーパスという新しい流れが生まれた。1990年、Sylviane Granger(ベルギー、ルーバン・カトリック大学)によって立ち上げられた International Corpus of Learner English (ICLE)がそれである。これは英語を外国語として学ぶ各国の大学生が書いた英語を系統的に収集する大規模なプロジェクトである。

ICLE の特徴はその系統的なコーパス構築の手法にある。ジャンルは argumentative essay で、長さは500語以上1000語以下と定め、第一言語別にそれぞれ20万語のサブコーパスが作られている。また、母語話者の英語と比較するため、ICLEと同じフォーマットで英米人大学生のデータも収集されている。日本では昭和女子大学の金子朝子氏が中心となってデータ収集が続けられている。

第二言語習得研究以外の目的では Longman 社が学習

辞書編纂のため作成した Longman Learners' Corpus がある。これは各国の学生のエッセイや試験の解答を集めたもので、1 千万語の規模がある。1995 年に刊行された LDOCE 第 3 版では早くもその成果が利用されている。

シンポジウム (2)

《コーパスを利用した英語教育と英語・英文学研究指導 実践報告と今後の可能性》

(司会 八木 克正)

このシンポジウムは、コーパスを英語教育と英語・英文学の研究指導にどのように生かすかという観点から、今までの実践と今後の可能性について意見交換をし、理解を深めて行く場としたい。今、小中高大のすべての学校での教育のあり方と質が問われている。その時代にあつて、コーパスあるいはコーパス言語学が、英語教育、英語・英文学研究指導のあり方にどのようなインパクトを与えてきたか、また今後どのような役割を期待できるのか、という見通しを立てるきっかけを作りたいと思う。

今後の英語教育や英語・英文学の研究指導の中で、コーパスは文字通り欠かせないものになってゆくであろう。しかし今の所、教育・研究指導の中で、コーパスがどういう役割を果たすことができるのか、どのような可能性を持っているのかについて、必ずしも共通の理解があるわけではない。そこで、それぞれの分野で豊富な研究実績があり、かつ教育・研究指導の経験豊かな講師にご出席をお願いした。

まず、各講師から、「英語教育」「英語研究(共時的視点から)」「英語研究(通時的視点から)」「文学研究」それぞれの分野について、今までの実践と今後の展望についての考えを聞くことにする。そしてその後、フロアの方々と討論を通じて、今後の「教育と研究指導」にコーパスをどのように利用してゆくか、どのようなコーパスが必要か、コーパス言語学的な視点がどのような役割を果たすことができるのかなどについて、考えを深める機会にしたい。

コーパスを英語教育に生かす

(投野 由紀夫)

コーパス言語学の言語教育への応用は興味ある研究課題である。この発表では短い時間ではあるが、世界的な動向をまとめた上で、日本の英語教育ではどのような

可能性や課題があるかを整理してみたい。

1) コーパスの教育利用：世界の動向

コーパスの教育への利用の先駆は COBUILD project である。またバーミンガム大学の Tim Jones を中心とした data-driven learning の考え方は今でもヨーロッパで非常に根強い支持者がある。バーミンガム大学の路線とは少し異なる方向性でランカスター大学で始まった Teaching and Language Corpora (TALC) という隔年の学会があり、ヨーロッパ全体に活動が広がって今年で 5 回目を迎える。この学会の特色は、コーパスを方法論として位置づけ、teaching に関しても幅広い定義を受け入れている点である。北米では AAACL という学会が 4 回目の大会をこの 11 月に開催する。「応用コーパス言語学(applied corpus linguistics)」と称して、コーパス言語学の知見を各応用分野に生かすという米国らしい発想がある。これらの学会に参加した経験をアジアの動向なども含めて簡単に整理する。

2) コーパス言語学の英語教育への利用

コーパス言語学の英語教育での利用は大別すると、以下のような領域になる。発表では各領域に関してコーパス利用の可能性と今後の課題を以下のようにまとめる。

- 教授資料 (teacher reference) : コーパス・データを教材研究などに活用するためには?
- 教材作成 (materials design) : 辞典以外のコーパス活用の方法は?
- シラバス作成 (syllabus design) : コーパス・データの有効性は?
- 教師教育 (teacher training) : コーパス・データは教員の言語観を変えられるか?
- e-learning : CALL システムの一部としてコーパスの果たす役割は?
- テスト作成 (testing) : コーパスを用いるとテスト作成や評価はどう変わるか?

最後に、これらの領域の展望を踏まえて、英語コーパス学会のような専門家集団の果たすべき役割や課題を議論したい。

コーパスを現代英語研究の共時的な研究指導に生かす

(梅咲 敦子)

コーパスを現代英語の研究指導に生かす方法を考える場合、コーパスを現代英語研究にどのように利用して

何を導き出すかを中心とした言語学的視点と、言語分析のうち何をどのように扱うことが学生にとって有効かを中心とした教育学的視点が必要になる。言語学的には、コーパスを利用して、音声・音韻、語彙・文法、ディスコースレベルの分析を行い、既存の説を検証したり実証的に新たな発見をすることをめざすが、教育的には、学生が自らが達成感を得るための指導を目標に、学生のニーズや教育環境を考慮してシラバスを作成し実践する必要があると考えられる。

本発表では、まず、実践の一例として、発表者がこれまでに、短期大学、四年制大学および大学院で担当してきた科目「英語情報処理演習」「英語情報処理論」「コーパス言語学特殊講義」の教育目標、実践の概要と結果を紹介する。そのなかでは、コーパスとして、主に BoE (オンライン) および Brown, LOB, Frown, FLOB (ICAME Corpus Collections on CD-ROM), また学生が作成したミニコーパスを用いて、語彙リストやコンコードスライン、また統計データを、コーパス付属検索ソフトや KWIC Concordance for Windows などのソフトを用いて出力した。また、近年は、Web ページを利用したコーパス検索、電子テキストの入手、タグ付けなども授業実践に取り入れ、得られた結果から、綴り字と語彙の地域差、語法、テキストの特徴について考察を加えてきた。

次に、これまでは、文章英語を中心とした語彙・文法レベルの現代英語研究指導が中心であったが、今後の可能性として、例えば、SEC のような音声版つきコーパスを利用した韻律特性の研究とその指導例について触れる。また、話し言葉の研究指導の実践例と可能性について、Michigan Corpus of Academic Spoken English (MICASE) や Santa Barbara Corpus of Spoken American English の利用にも触れる。さらに、広くレジスターごとの英語の特徴を研究するために、特殊コーパス編纂を授業に導入する可能性についても考えたい。

コーパスを英語の通時的な研究指導に生かす

(大門 正幸)

語の歴史的変化を広い視点から理解するためには、音の変化、形態の変化、統語構造の変化、意味の変化といった英語の言語的な変化(内面史)だけでなくこれらの変化を取り巻く政治・社会・文化的要因(外面史)にも目を向ける必要がある。本発表では、英語の内面史および外面史について教育・研究指導していく場合、どのよ

うな形でコーパスが利用可能であるかについて考察する。

外面史に関しては、重要な出来事に関する一次資料の提示、つまり資料集としてコーパスを利用することが可能である(ただしこの場合には「コーパス」という語を広義に解釈する必要がある)。たとえば、theatre と theater の対立に見られるようなイギリス英語とアメリカ英語のつづり方の相違がウェブスターを代表とする改革論者達の活動に由来することを指導する場合、Historical Resources などを利用して実際のウェブスターの主張に触れさせることが考えられる。発表者がこれまで英語史の教育・研究指導において実践してきたのは主にこのような(広義の)コーパスの利用である。

内面史に関しては、各時代の代表的なテキストで作成したコーパスをもとに、英語史における重要な変化を簡単に提示できるようにする。統語的变化を例にとれば、OV 語順の衰退と VO 語順の発達という変化を扱う場合、当該の語順が各時代のテキストにどれくらい存在するのかを瞬時に検索し、それぞれの時代における当該の語順の数(頻度)グラフなどで視覚的に分かりやすくしたものを提示するというような利用法が考えられる。このような場合、統語的標識付けをされたコーパスを用いることができる。古英語と中英語においては統語的標識が付けられた Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English や Penn-Helsinki Parsed Corpus of Middle English を利用することが可能であるが、近代英語期以降についての変化を扱うためにはこれらのコーパスと同じ方法で標識付けされたコーパスの作成が必要である。また、特定の単語を検索することでそれぞれのテキストにおける当該の単語の語形・意味を検索することができるようなコーパスがあれば、屈折の消失という形態論的变化、ノルマン人の征服によってもたらされた正書法の変化、語の意味変化などについても実際の用例に基づいた説得力のある指導が可能となるであろう。コーパスを用いて内面史の指導を行うにはこのような視点に立って作成されたコーパスの作成が必要である。

コーパスを英文学研究指導に生かす

(石川 慎一郎)

英文学研究においては、19 世紀以来、長らく印象批評的な方法論が主流となってきた。印象批評は、幅広い読書経験を持った批評家によって適切に行われる場合

は、深い洞察や気付きを含む説得力のあるものになりうるが、そうでない場合は、根拠のない感想の羅列に墮する危険性もはらんでいる。

20世紀に入って、文学研究が大学のカリキュラムとして組織化され、また、大学自体の大衆化が急速に進むと、文学研究には、明証的に指導できる実証的な科学としての側面が強く求められるようになった。こうして、とくに戦後の米国の英文学研究現場を席卷したのが、いわゆる「新批評 (new criticism)」である。新批評は、作品を作者から切り離して「テキスト」と呼び、わずかな量のサンプルテキストを、純粋な言語構築物として徹底的に分析することを主たる目的とする。

しかしながら、作品の概念をあまりに狭く捉えることと、一部で全体を語ろうとすることが限界と目されるようになり、新批評は次第に影響力を失っていった。その後の文学研究は、深層心理学的批評・構造主義批評・脱構築批評・ポストコロニアリズム批評・カルチュラルスタディーズといった多様で難解な文学理論が転変する一種の混沌の状態にある。こうした状況の中で、大学における(カリキュラムとしての)英文学研究は、再び、実証的・科学的な方法論を強く求めているのである。この要請にもっとも叶うと思われる学問分野の一つが「文体論(stylistics)」とくに「言語学的文体論 (linguistic stylistics)」である。

その提唱者と目される G. Leech and M. Short, *Style in Fiction* (1980) に寄せた序文の中で、Randolph Quirk は、「文体論」という新しい分野においては「言語学者も文芸批評家も、究極的には別物である互いの組合員カードを見せることなく、また、らちもない境界論争に明け暮れることもなく、協力して仕事を進めることになるだろう」(リーチ・ショート『小説の文体』: 笈壽雄監訳・瀬良晴子・廣野由美子・石川慎一郎共訳: 近刊予定)と将来を予言してみせた。しかしながら、その後約20年を経て、Quirkの予言した言語学と文学の融合は思ったほど実現していないのが実情である。

その理由のいくらかは、科学を志向する上で文体論が重視した計量という手法が、文学テキストという膨大なサンプルに適用されるには、あまりにも消耗的であり、文学研究者が、計量を終えて、さらにその先の真に意味のある分析の段階に進むことが困難であった点にある。この意味において、計量の徒労から研究者を解放つコーパス処理という手法は、文体論を再び活性化し、とくに大学教育における英文学研究の主たる方法論として

再生する有効な装置となりうると思われる。

特別講演

Corpus Linguistics: Past, Present, Future

(Stig Johansson)

It is now about forty years since the first computer corpus was compiled, the well-known Brown Corpus. We have come a long way since then. In looking back I will touch on the history of ICAME, the International Computer Archive of Modern English (now: the International Computer Archive of Modern and Medieval English). I will comment briefly on my work on the LOB Corpus, the Lancaster-Oslo/Bergen Corpus. I will show how the interest in corpora spread and how the work developed in quantity, breadth, and sophistication.

In talking about the present I will touch on my work on the Longman Grammar of Spoken and Written English. This is a grammar which describes language in use, with reference to a large corpus of different types of texts. The focus is on the analysis of quantitative patterns in four registers: conversation, fiction, news reportage, and academic prose. I will give some examples from the studies I undertook and comment on some difficulties I encountered.

One of the most promising developments in recent years is the compilation and analysis of multilingual corpora for use in cross-linguistic research. I will comment on problems and prospects in this area, with special reference to our work on the English-Norwegian Parallel Corpus (www.hf.uio.no/iba/prosjekt) and the Oslo Multilingual Corpus (www.hf.uio.no/german/sprik). With the aid of multilingual corpora we get unprecedented opportunities to study and contrast languages in use, including frequency distributions and stylistic preferences. Such corpora are absolutely essential for discourse studies, but they will also enrich studies of lexical and grammatical patterns. Applications include the production of new bilingual dictionaries and contrastive grammars and the development of materials for foreign-language teaching and the training of translators.

Finally, I will try to sum up the prospects of corpus research. After close to forty years of corpus-based language research, much has been achieved, but it is likely that we are only at the beginning of the era of corpus linguistics.