

英語コーパス学会 第30回記念大会資料

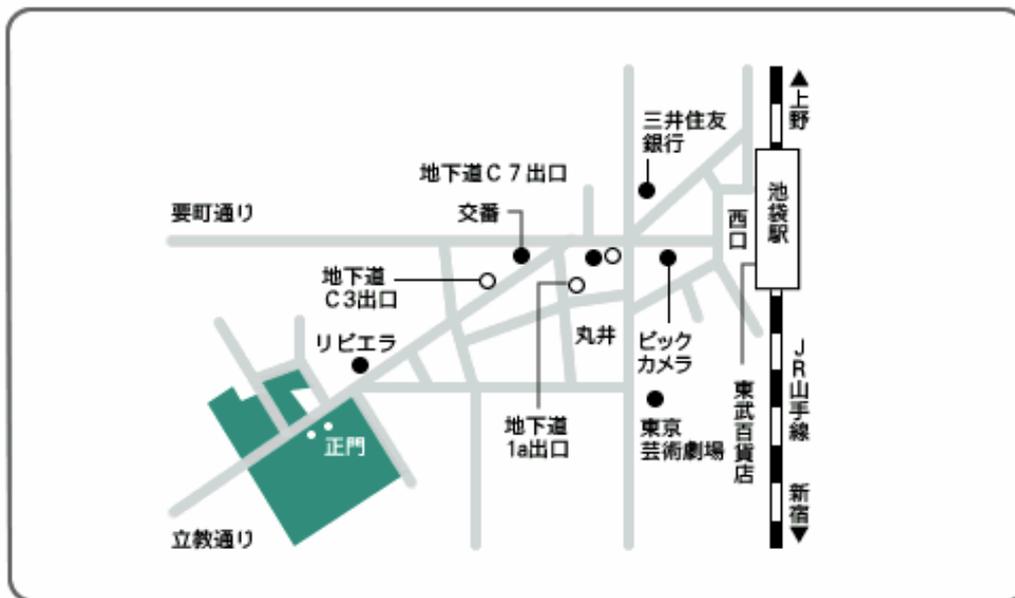
日時: 2007年10月6日(土)午後1時より(正午受付開始)

会場: 立教大学池袋キャンパス 7号館1階

(<http://www.rikkyo.ne.jp/>)

〒171-8501 東京都豊島区西池袋 3-34-1

後援 立教大学英語教育研究所



JP山手線・埼京線・高崎線・東北本線、東武東上線、西武池袋線、地下鉄丸ノ内線・有楽町線「池袋駅」下車。西口より徒歩7分。

英語コーパス学会 2007 秋

第 30 回記念大会プログラム

ワークショップ《 R を用いたコーパスデータの統計解析 》

会場：8号館 4階(メディアセンター) 時間：10:00～12:00(9:30 受付開始)

講師：金 明哲(同志社大学)

定員：先着 60 名

参加費：会員無料・非会員 1,000 円(申し込みは郵送・電子メールで事務局まで)

日 時 2007 年 10 月 6 日(土)
会 場 立教大学池袋キャンパス 7号館 1階
受付開始 12:00(7号館 1階)
開 会 13:00

司 会 山崎 俊次(大東文化大学)
中村 純作(立命館大学)
服部 正治(立教大学総長補佐)

1. 会長挨拶
2. 開催校挨拶
3. 事務局からの連絡

< 研究発表第 1 室(7101) >

研究発表 1 13:30～14:00 司 会 塚本 聡(日本大学)
古英語における語順決定要因の解明とデータベースの構築

鈴木 敬了(大東文化大学)

研究発表 2 14:05～14:35 司 会 瀬良 晴子(兵庫県立大学)
Kennedy 以降のアメリカ大統領による一般教書演説からの特徴語の抽出

大羽 良(早稲田大学非常勤)

研究発表 3 14:40～15:10

コーパスと文学的読みの融合：ヘミングウェイの短編を例にして

堀 正広(熊本学園大学)

< 研究発表第 2 室(7102) >

司 会 加野 まきみ(京都産業大学)

研究発表 1 13:30～14:00

OEDs、JUICE 中の日本語からの借用語、貸出語の形式的・意味的特徴の量的分析

藤原 康弘(中国学園大学非常勤講師)

研究発表 2 14:05～14:35

英語学習者の使用語彙と発話の通じやすさの関係に関する考察：誤り語と訂正語の意味的関連性に基いて
和泉 絵美(独立行政法人情報通信研究機構)
内元 清貴(独立行政法人情報通信研究機構)
井佐原 均(独立行政法人情報通信研究機構)

< 休 憩 15:10～15:30 >

シンポジウム 15:30～18:10(7101)

《 他言語コーパス研究の現在：英語研究への示唆 》

司 会 滝沢 直宏(名古屋大学)

コーパスによる自発音声の韻律特徴の分析

講 師 前川 喜久雄(国立国語研究所)

フィンランド語記述文法とコーパスデータの役割

講 師 千葉 庄寿(麗澤大学)

タガログ語データ/コーパスの質と性格

講 師 大和田 栄(東京成徳短期大学)

フランスの特徴的なコーパス研究：語彙研究と政治ディスコース研究

講 師 藤村 逸子(名古屋大学)

閉会の辞

鳥飼 慎一郎(立教大学)

《 懇親会 時間：18:30～20:30 場所：太刀川記念館 3階多目的ホール 会費：5,000 円 》

【ワークショップ】

R を用いたコーパスデータの統計解析

金 明哲(同志社大学)

コーパスデータを科学的に解析するためには、統計学の基本知識とデータ解析のツール・プログラムが不可欠です。データ解析のツール・プログラムはすでに多く存在し、毎日のように新しい研究開発が行われています。市販されているデータ解析の専用ツールとしては、SAS、SPSS、S-PLUSなどが広く知られています。しかし、機能の複雑さ、頻繁に行われるバージョンアップなどに対応するのに悩まされている研究者は少なくありません。このようなことからフリーソフト R の普及が急速に進んでいます。R 上で使用可能なパッケージは、すでに 1,000 を大きく上回っています。そこで、本ワークショップでは R をツールとして、コーパスのデータ解析について実例を用いて学習・演習を行います。ワークショップでは、R の基本操作やコーパスの統計データの出入力から始め、探索的データ解析の技法(基本統計量、データの視覚化、データの縮約とクラスタリング等)、推測統計(t 検定、カイ二乗検定等)とデータのモデリング(線形、非線形モデル等)の技法について、参加者のご要望に合わせて説明する予定です。ご要望は mjin@mail.doshisha.ac.jp にお問い合わせいたします。

【研究発表第 1 室】

【研究発表 1】

古英語詩における語順決定要因の解明とデータベースの構築

鈴木 敬了(大東文化大学)

古英語詩に関する語順研究は数少ないが、それらにおいても韻律の影響は論じられていない。古英語散文における法助動詞の語順研究に関して Ohkado (2000) では 'extra element' (主語、法助動詞(M)、本動詞(V)以外の要素)の有無が MV/VM の語順決定要因であるとしているが、古英語詩 Beowulf においては 'extra element' の有無にかかわらず VM 語順が多く、その作用は見られない。散文における語順決定要因として 'heaviness' (要素の長さ)の作用を多くの研究者(Kohonen 1978、Suzuki 1993a、Davis 1997 など)が主張しているが、V の 'heaviness' は要因とはならず、韻律上の M の 'heaviness' の影響が限定的に見られるのみである。以上、散文での要因は詩では見られないことから、Suzuki (2004) では詩の特徴である頭韻パターンと語順の関係を調査し、頭韻の分布が法助動詞の重要な語順決定要因であると主張した。

本発表では、こうした研究を踏まえ、古英語散文から韻文に翻訳されたとされる Boethius の作品を資料とし、ジャンルによる語順決定要因の違いをより明確にし、従来の口承定型句理論の有効性を含め、詩人がどのように作詩したかについて考察したい。同時に現在の歴史的コーパスには詩形の情報が含まれていないことから詩の語順研究のためのデータベース構築を提案したい。(本発表は今年度科研費補助金 基盤研究(C) 課題番号 17520339 による研究の一部である。)

【研究発表 2】

Kennedy 以降のアメリカ大統領による一般教書演説からの特徴語の抽出

大羽 良(早稲田大学非常勤)

本発表は第二次世界大戦後の新しい時代を予感させたケネディ以降から現代までの 9 人の大統領が

行った一般教書演説から、コーパスを作成し、各大統領のサブコーパス中で用いられる特徴語を複数の統計的指標を用いて同定することを目標とする。また一般教書演説という政治的言説の中に現れる特徴語が大統領の語の使用のスタイルだけでなく、施政方針やその当時の社会的状況・彼らの思想性とどのように関連付けが可能かを考察する。最後に特徴語を同定するために使用した統計的指標から得られた数値をもとに、多変量解析の手法を用いて分析し、大統領の教書演説コーパスと特徴語がどのように分類されるかを見る。研究方法としては、最初にウェブサイト上から収集したデータをもとに品詞・文境界などのタグを付けた一般教書コーパスを作成し、基本語彙統計量を求めコーパスの特徴を概観したのち、中條他(2005)で扱われた統計的指標をもとに複数の指標を用いて各大統領の教書サブコーパスから特徴語を抽出する。そして各指標の上位に位置づけられた語に対して質的な分析を行う。次に各大統領の一般教書サブコーパス間で特徴語の重なり度を計算し、それをもとにそれぞれの統計的指標の特徴を概観する。最後に一般教書コーパス中に高頻度で現れる上位 100 語を選び、それぞれの大統領のサブコーパスから得られる各語の統計的指標を用い、多変量解析のひとつである主成分分析によって各大統領サブコーパスがどのように分類されるかをみる。このように様々な統計的指標を用いることでよりテキストの特徴を捉えることが可能であることを論じる。

【研究発表 3】

コーパスと文学的読みの融合：ヘミングウェイの短編を例にして

堀 正広(熊本学園大学)

文学作品の言語文体研究へのコンピュータ利用やコーパス作成による分析は新しい研究領域として注目されるようになった。2005 年 Birmingham 大学で開催された Corpus Linguistics 2005 では、pre-conference workshop として“Corpus approaches to the language of literature”が行われ、2006 年 Nottingham 大学で開催された The Third Inter-Varietal Applied Corpus Studies では Martin Wynne と Peter Stockwell による“Corpus Stylistics: a Public Inquiry”という発表が行われた。また、2006 年と 2007 年には英国に本部を置く国際文体論学会(PALA)の pre-conference でも同様の workshop が行われた。

文体研究へのコーパスからのアプローチは Corpus Stylistics という用語で表現されるようになり、*Corpus Stylistics* (E. Semino and M. Short, 2004)が出版され、また、*Approaches to Corpus Stylistics: The Corpus, the Computer and the Study of Literature* (D. Hoover, J. Culpeper, and M. Wynne, 2007)も刊行予定である。しかし、文学言語へのコーパスの利用はその理論や方法論、客観性と解釈等の点でまだいろいろと検討の余地があるように思われる。

従って、本発表では、文学作品の言語文体研究におけるコーパス利用と文学的な読みとの融合、つまり、コーパス利用によって得られたデータとそのデータを分析する言語理論、そして文脈的な解釈に基づいたバランスのとれた文体の分析法を提起したい。例として、ヘミングウェイの短編にみられる男女の齟齬、つまり心のすれ違いは、会話のスタイルの違いに見られることを、機能語の頻度やコロケーションに注目して文脈を考慮しながら明らかにしていく。理論的な枠組みとして Michael Hoey の *Lexical Priming* (2004) と Deborah Tannen の *Conversational Style* (2005 New Edition) を用いる。

【研究発表第 2 室】

【研究発表 1】

OEDs, JUCE 中の日本語からの借用語、貸出語の形式的・意味的特徴の量的分析

藤原 康弘(中国学園大学非常勤講師)

本発表の主たる目的は、5 種のオックスフォード系英語辞書(OEDs)の日本語からの「借用語」と、

日本人英語使用者コーパス(JUCE、藤原 2006、2007)にみられる日本語からの英語への「貸出語」の形式的・意味的特徴を量的に分析し、その結果から今後英語に借用される可能性の高い語彙を提示することである。

近年、言語学、とりわけ諸英語(Kachru 1986、1992)、国際英語(e.g. Smith 1983; Crystal 2003)の観点から、言語接触における明示的な現象として英語の借用語への関心が高い。具体的には、1) 英語母語圏で発刊ないし編纂されている英字新聞・辞書・コーパスなどにおける他言語、とりわけドイツ語(Stubbs 2002)と日本語(e.g. Kimura-Kano 2006; 早川 2006)からの借用語、また 2) インド(Dubey 1991)、中国(Yuan 2005)、ハワイ(島田 2006)などの非英語母語圏で発刊された英字新聞に見られるヒンドゥー語、中国語、日本語からの借用語の研究が進められてきている。上記の出处を異にする2種類の借用語を、理解の便宜上、前者を「借用語」(borrowing)、後者を「貸出語」(loan)と区別し、OEDsの借用語、JUCEの貸出語を、近年に出版された最大の日本語意味分類辞書『分類語彙票：増補改訂版』(2004)を1つの客観的な指標として、形式・意味範疇による分析を行う。その得られた結果にPearson 2乗検定・尤度比検定を行い、今後どのような日本語が英語へ借用語として認知される可能性があるか示したい。

【研究発表2】

英語学習者の使用語彙と発話の通じやすさの関係に関する考察：誤り語と訂正語の意味的関連性に基いて

和泉 絵美(独立行政法人情報通信研究機構)

内元 清貴(独立行政法人情報通信研究機構)

井佐原 均(独立行政法人情報通信研究機構)

発表者はこれまで、非母語での発話に含まれる誤りがその発話の通じやすさに及ぼす影響について、日本人英語学習者の話し言葉データを用いて分析を行ってきた。その結果、誤りの種類によって影響の度合いが異なり、特に、発話の意味内容を担う語彙・談話に関する誤りが発話を通じにくくする主な原因となっていることがわかった。本研究では、このうち名詞の語彙誤りが及ぼす影響について分析を深める。

まず、データ内の名詞(単語レベル)に付与されているエラータグ(誤りの種類と訂正語)から、誤り語と訂正語のペアを得る。次に、WordNetで記述されている概念階層を用いて、各語がその文脈においてどの語義で用いられているか特定し、2語の語義の意味的関連性を求める。意味的関連性(semantic relatedness)とは、carとbicycleのような類似性のほか、carとgasolineに見られるような何らかの意味的な繋がりのことである。これまでに提案されているいくつかの尺度を用いて、誤り語と訂正語の意味的関連性を求め、日本人英語学習者の語彙誤りの特徴を量・質の両面から分析する。量的な分析としては、意味的関連性を示す値が、発話の通じやすさレベル(intelligible, unclear, unintelligibleの3レベル)の違いや英語運用能力レベルの違いによってどのように推移するかを測定する。質的な分析としては、個別の事例から概念階層における誤り語と訂正語の関係を観察するほか、概念階層からの情報だけでは測りきれない事例についても考察を行う。最後に、本分析から得た知見の語彙指導改善への活用について検討する

【シンポジウム】

他言語コーパス研究の現在：英語研究への示唆

司会 滝沢 直宏(名古屋大学)

英語の世界におけるコーパスの利用は1960年代のいわゆるBrown Corpusの構築以来、ほぼ半世紀にわたる歴史があり、辞書の編集、語法文法研究、英語教育をはじめとする多方面における利用が既に

定着し、方法論的な蓄積もなされてきていると言ってよい。

勿論コーパスは、英語の世界においてのみ用いられているのではなく、他の多くの言語においても用いられている。しかるに、英語以外の言語におけるコーパスの利用状況やコーパスを用いた言語研究の手法や関心の持ち方などに対して、英語研究者が注意を払い、そこから何かを学び取るという試みはほとんど行われていないように思われる。

本シンポジウムでは、このような現状を考慮し、また言語によって拠って立つ理論や研究の視点、そして研究手法などが異なることを踏まえ、敢えて英語以外の言語を研究している方々に発表して頂き、その中から英語研究に対してどのような示唆を得ることができるのかを考える。

コーパスによる自発音声の韻律特徴の分析

講師 前川 喜久雄(国立国語研究所)

音声の韻律的特徴の研究はデジタル音声処理技術の実用化と音韻理論における理論構築とに支えられて 20 世紀最後の四半世紀にめざましい発展を遂げた。しかしそこで検討されたのは、実験面においても理論面においても、いわゆる朗読音声であった。音声の韻律的特徴が語の意味よりはむしろ句や発話の意味に関与しており、自立的であるよりもむしろ状況依存的事であることを考えると、韻律的特徴の研究サイクルは理論と実験だけで完結するものではなく、理論ないし実験研究の成果を大量の自発音声データに照らして検討することが必要であると考えられる。従来、世界のどこでもそのような研究が実施されてきていないのは、単に研究のインフラが整備されていないことの結果である。

筆者らが開発した『日本語話し言葉コーパス』(CSJ)には大量(660 時間以上)の自発音声格納されている。発話は形態素解析を含む豊富なアノテーションが施されているほか、約 50 万語分については韻律的特徴の精密なラベル情報が提供されているなど、自発音声研究用コーパスとして、世界的にみて画期的な仕様を備えている。

今回の発表では筆者がこれまでに CSJ を利用して実施した日本語の韻律的特徴に関する研究の成果をかいつまんで報告する。以下のようなトピックに触れる予定である。1) 2 モーラ有核助詞のフレーズング、2) 韻律境界強度(Break index)の推定、3) L%HL%句境界音調(Boundary pitch movement)の変異。

フィンランド語記述文法とコーパスデータの役割

講師 千葉 庄寿(麗澤大学)

フィンランド語はウラル語族のバルト・フィン諸語に属する言語である。語形変化の少ない英語と比較すると、フィンランド語は膠着語的な性質が強い。例えば、1つの名詞には、10を超える格のほか数、所有接辞、多様な小辞が接続し、屈折変化だけで2,000を超える語形が存在する。このような類型論的特徴をとらえるため、コーパスを用いたフィンランド語の研究では正規表現や形態素解析技術が駆使される。

80年代前半、音韻論・形態論・統語論の各分野でコーパスから得られた量的データを用いた研究がおこなわれたことが契機となり、フィンランド語研究においてコーパスは比較的早い段階から重要視されてきた。最近公開された大規模なフィンランド語記述文法 *Iso suomen kielioppi*(2004) にはコーパス分析の成果も大々的に取り入れられ、ここではコーパスデータは記述の精密化だけではなく、これまでのフィンランド語学の知見の整理と体系化にも貢献している。

本発表では、フィンランド語コーパスとその分析手法について現状を紹介するとともに、フィンランド語記述文法の発展にコーパスデータが果たしている役割を検証し、特に、コーパスの規模の変化が文法記述にどのような変化をもたらしたかを最近の成果を踏まえ考察する。最後に、フィンランド語研究における大規模コーパス利用の現状に見える課題を踏まえ、言語の類型を意識したコーパス分析の枠組みづくりの必要性について、フィンランド語研究の立場からコメントしたい。

タガログ語データ / コーパスの質と性格

講師 大和田 栄(東京成徳短期大学)

タガログ語は、フィリピンにある100以上の諸言語の中で、国内ではかなり支配的な位置を占める言語の一つであり、新聞・雑誌を含めた出版物やWebサイトなどの数も圧倒的に多い。それをコーパスとして利用すること自体は、他のフィリピン諸語に比べるとかなり容易であるが、きちんと手をかけられ、一般に使用可能で大規模なコーパスは存在しないようである。また、フィリピン語は、1987年の憲法の下で、英語とともに公用語の位置を占めている言語で、研究者の中でも捉え方が色々あるが、タガログ語をベースにした言語という把握の仕方にそれ程大きな間違いはない。2言語が公用語であることから、憲法を始め、公文書には両言語が使われ、英語・フィリピン語の平行コーパスとして使用できる可能性もある。一方、タガログ語(フィリピン語)には、スペイン語・英語からの借用語も多く、コードミキシング(英語・タガログ語間)も、話し言葉だけではなく、書き言葉においても観察される。

本発表では、タガログ語について一般的に言及される語順(VSO/VOS)や「焦点格/話題格」などと呼ばれるものなどについて、いくつかの種類のデータから見られる事実の相違などについて概観したい。

フランスの特徴的なコーパス研究：語彙研究と政治ディスコース研究

講師 藤村 逸子(名古屋大学)

フランス語で Corpus は「資料体」という意味で、人文科学、社会科学、哲学、法学などの分野で用いられてきた伝統的な用語である。現代の言語科学においても、コーパスとは電子化された巨大なデータを必ずしもイメージさせないこと、またコーパスの構築・選択、およびコーパスから得られたデータの解釈に重大な関心が払われることは、この伝統に由来する。フランスの特徴的なコーパス研究として、二つを挙げたい。これらの二つは対照的ではあるが、意味に関心の中心がおかれたものであるという点で共通性がある。一つは、語彙研究とそれに付随する計量言語学的研究である。辞書: Trésor de la langue française (10万語、19-20世紀)の編纂とそのための電子コーパス: Frantext (2億語以上、中世フランス語から現代まで)がその代表とされる。1950年代終わりに始まったこのプロジェクトは、1950-60年代の機械翻訳の興隆がきっかけとなったもので、その当時、多くの言語学者が自然言語処理に関心を持った。言語学と情報工学の結びつきは強く、同じ研究組織に所属する言語学と情報工学の専門家が共同研究する状況は広範に見られる。もう一つは、ディスコース分析の一分野としての、政治ディスコースの研究であり、政治家の演説を集めたコーパスが収集・公開されている。この分野では、コーパスの構築が重要課題であり、コーパスは比較的小規模であるために、質的分析が中心であったが、最近では計量的分析方法も盛んになっている。このように、「パロール」の言語研究と「ラング」の言語研究を結びつける努力が全体としてはなされていると思われる。

《大会参加者へのご案内》

- 自家用車でのご来場はできません。
- ワークショップの受付は「8号館4階(メディアセンター)」で午前9時30分から行います。
- 大会の受付は「7号館」1階7101教室前で正午から行います。
- 昼食については、大学食堂、一般利用食堂、専門店街などが利用できます。
- 校内は分煙措置がとられています。指定場所での喫煙にご協力ください。
- 会員でない方も、「当日会員」として参加していただけます(会費 1,000円)。

キャンパスマップ



2007年8月31日 発行
編集・発行 英語コーパス学会
代表者 中村 純作
事務局 〒175-8571 東京都板橋区高島平 1-9-1
大東文化大学 山崎俊次研究室内
TEL: 03-5399-7372 FAX: 03-5399-7373
E-mail: yamazaki@ic.daito.ac.jp
URL: <http://muse.doshisha.ac.jp/JAECS/index.html>
