

英語コーパス学会 Newsletter No. 68

Mar. 1, 2010

■会長:赤野 一郎
■事務局:〒770-8502 徳島県徳島市南常三島町 1-1 徳島大学総合科学部 井上永幸研究室内
■TEL:088-656-7125 ■振替口座:00940-5-250586(英語コーパス学会)
■URL: <http://muse.doshisha.ac.jp/JAECS/index.html> ■email: naga.inoue@gmail.com

JAECS
Japan Association for English Corpus Studies

第 35 回大会のご案内

英語コーパス学会第 35 回大会は、4 月 24 日(土)、兵庫県立大学(神戸学園都市キャンパス; 〒651-2197 神戸市西区学園西町 8 丁目 2-1; <http://www.u-hyogo.ac.jp/>)で開催されます。会場となる神戸学園都市キャンパスは、神戸市営地下鉄学園都市駅下車徒歩約 15 分のところにあります。事前に経路と時刻をご確認ください。宿泊を必要とされる方は早めの予約が肝要かと思われます。なお会場校の責任者である瀬良晴子先生には準備段階からご尽力いただき深く感謝致します。

詳細は、同封の「大会資料」をご覧いただきたいのですが、恒例になりました午前中のワークショップのほか、午後には研究発表 4 件と講演会、そしてシンポジウムがあります。

午前中のワークショップでは、投野由紀夫氏(東京外国語大学)に、「Web コーパス検索ツール Sketch Engine の基本操作と活用」と題して、世界的に注目を集めている web コーパス検索ツール Sketch Engine の概要と基本操作、目的に応じた活用方法をわかりやすく紹介していただきます。

研究発表については、運営委員の査読を経て準備委員会が最終的な審査を行った結果、次の 4 件の発表が決まりました。鈴木大介氏(京都大学大学院生)の「no doubt の機能について 一類義表現との比較から」、神谷昌明氏(豊田工業高等専門学校)の「古英語・中英語に現れる小節・結果構文 一通時的英語コーパス YCOE と PPCME 2 を検索して」、宇佐美裕子氏(ランカスター大学大学院生)の「コーパスを応用した言語テスト問題の改良 一日本の大学入試問題から文法 4 択問題を例に」、石田知美氏(名古屋大学大学院生)・杉浦正利氏(名古屋大学)の「日本語母語英語学習者の連語表現の判別分析」で、2 室に分かれての研究発表となります。

講演会ですが、大名力氏(名古屋大学)に、「コーパス検索で注意すべきこと 一基礎データの信頼性向上のために」と題して、基礎データである出現頻度の算出に焦点を当て、具体例を挙げながら、コーパス検索やデータの分類・解釈において注意すべき点について講演いただきます。

シンポジウムは、投野由紀夫氏(東京外国語大学)

の司会の下、「進化する Web コーパス:現状と課題」と題して、めざましく進化を続ける web コーパスの現状を把握し、英語コーパス研究における web コーパスの位置づけや役割、また問題点や課題などについて、web コーパスを積極的に研究・活用している諸氏を講師に招いて、学会員に有益な情報を提供するとともに、一緒に web コーパスの意義や活用法などに関して活発な議論をしていただきます。まず、司会の投野由紀夫氏に「Web コーパス概観」と題して発表いただき、次に、田中省作氏(立命館大学准教授)が「Web コーパスの言語情報処理基盤」、大羽良氏(中央大学助教)が「Web コーパス研究におけるブログの可能性とその文体的特徴」、中村隆宏氏(ネットアドバンス SE)・星野守氏(小学館コミュニケーション編集局プロデューサー)が「小学館 Web コーパス“Sekai Corpus”構築とその活用」のタイトルでそれぞれ発表されます。

以上のように、今回は、Web コーパス検索ツール Sketch Engine に関するワークショップ、それに語法研究、英語史、言語テスト、学習者コーパスといった 4 件の研究発表、基礎データの信頼性向上をテーマとした講演、さらに web コーパスの研究活用をテーマとしたシンポジウムと盛りだくさんの内容となっておりますので、多数のご参加を期待しております。

午前中のワークショップに参加ご希望の方は郵便・電子メール(件名「ワークショップ申込」、宛先は、今回はワークショップ専用の naga.inoue@gmail.com でお願いします)で、所属と会員・非会員の別を明記の上、事務局までお申し込みください。英語コーパス学会の会員であれば参加費は無料です(非会員の場合は当日会費 1,000 円)。

会誌『英語コーパス研究』第 17 号について

『英語コーパス研究』第 17 号(2010 年刊行)について、進捗状況を報告いたします。2009 年 9 月に投稿された論文の査読審査はすべて終了し、現在印刷工程に入っています。掲載論文は、研究論文 7 点(投稿数 12)、研究ノート 1 点(2)、シンポジウム論文 2 組(2)となりました。過去数年に比べ、多数の投稿がありましたので、今回は掲載論文数も多くなり、200

ページを超える大部なものとなる見込みです。査読にあたられた会員の方にはこの場を借りて感謝申し上げます。5月には印刷が完了し、次回ニューズレター発行時に合わせてお届けできる見込みです。

『英語コーパス研究』がお手元に届いた後、数ヵ月後には再び、投稿締め切りの期日が参ります。今回の投稿と同様、次回の会誌へも、多数の投稿を期待しております。

『英語コーパス研究』編集委員会委員長
塚本 聡 (日本大学)

第9回英語コーパス学会賞の応募締切迫る！

第9回英語コーパス学会賞を募集いたします。学会賞は英語コーパス学会の活性化のために設けられた賞です。自薦・他薦は問いませんので、是非、今年1年の研究活動を各分野で振り返っていただき、学会賞、奨励賞候補の研究をご推薦ください。

【対象】英語コーパス学会の目的にてらし、英語のコーパス言語学に関する優れた研究業績をあげた学会員(個人またはグループ)とする。ただし、奨励賞は英語のコーパス言語学に関する優れた研究業績(論文の場合、学会誌『英語コーパス研究』に掲載されたものに限る)をあげた35歳以下または大学院修了後の研究歴5年以下の学会員個人に限る。

【応募方法】自薦、他薦を問わない。

【提出書類】1) 同封の推薦理由書。2) 論文の場合は原物またはコピー。単行本の場合は事務局で用意するので送付は不要。

【提出先】事務局。

【応募期限】2010年3月31日(水)。

【発表】2010年度秋季大会。

学会賞選考委員会委員長
投野由紀夫(東京外国語大学)

JAECS 東支部活動報告

東支部では下半期にコーパス言語学の世界的研究者お二人をお招きして会を催しました。まず11月8日(日)大東文化大学において、Adam Kilgarriff氏をお招きして、コーパス検索システム Sketch Engine のツール講習会を実施しました(東支部と東京外国語大学 G-COE 共催)。約30名ほどの参加者があり、皆さん熱心に学んでいました。さらに12月8日(火)には大東文化大学で Douglas Biber 氏を招いて特別講演会を開催しました。大東文化大学の学生を含め計70名ほどが熱心に講義に耳を傾けていました。世界的に著名なコーパス言語学者お二人が同時期にいらしていただき、大変素晴らしい機会となりました。会場を提供してくださいました大東文化大学の山崎俊次先生、鈴木敬了先生にこの場をお借りして感謝申し上げます。

東支部支部長
投野由紀夫(東京外国語大学)

2010年度の大会日程と開催校

第35回大会 4月24日(土) 兵庫県立大学

第36回大会 10月9日(土) 東京大学(駒場キャンパス)

新入会員紹介(11月13日現在、Sは学生)

宇佐美裕子 ランカスター大学大学院 S
嶋津成子 ダラム大学院 S

事務局から

◇会費納入のお願い

2010年度会費(一般5,000円、学生3,000円)を、日本郵便にある払込取扱票を使ってお納めいただきますよう、ご協力をお願いいたします。日本郵便発行の受領証をもって領収書に代えさせていただきますので、ご了承ください。別途領収書が必要な方は、80円切手を同封の上、石川保茂(〒615-8558 京都市右京区西院笠目町6 京都外国語大学)までお申し出ください。払込取扱票の通信欄によるお申し出はご遠慮ください。

過年度会費未納の方は、2010年度分と併せてお納めください。会誌『英語コーパス研究』第17号は2009年度の会費を納入していただいた方にのみ、送付いたします。また、2年続けて会費未納の場合、Newsletterなどの送付を中止させていただきます。

住所、所属などに変更や異動のある方は、必ず払込取扱票の通信欄にお書き添えください。

※会員の皆様には、日頃より会費の当該年度内納入のご協力をいただきまして、お礼申し上げます。会費を滞納されますと、退会時に滞納分をまとめてお支払いいただくといった事態にもなりかねません。会員の皆様におかれましては、円滑な学会運営のためにご協力いただけましたら幸いです。

◇メーリングリストについて

英語コーパス学会ではメーリングリストを使って会員の皆様の様々な情報交換に役立てていただいているところですが、最近、宛先不明でエラーが返ってくる例も増えています。会員の皆様方には、メールアドレスに変更が生じた場合、速やかに事務局宛ご連絡いただけますようお願い申し上げます。



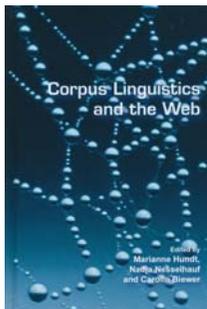
FORUM

■文献紹介

滝沢直宏 (名古屋大学)
takizawa@cc.nagoya-u.ac.jp

Hundt, Marianne, Nadja Nesselhauf
and Carolin Biewer (eds.) (2007)
Corpus Linguistics and the Web.
Amsterdam / New York: Rodopi.
305 pages.

US\$ 91 (hardcover)
ISBN-10: 90-420-2128-4
ISBN-13: 978-90-420-2128-0



本書は、2004年にドイツのハイデルベルクで開催されたシンポジウム Corpus Linguistics—Perspectives for the Future における発表論文の一部と若干の招待論文から構成された論文集である。編者の一人 Marianne Hundt は、現在、チューリッヒ大学に所属し、言語変異、史的統語論などを研究している。Nadja Nesselhauf はハイデルベルク大学に所属し、学習者コーパスを用いたコロケーション研究に関する著書がある。Carolin Biewer も同じくハイデルベルク大学の所属で、コーパス言語学や社会言語学を専門とし、現在、南太平洋の島々で使われている英語に関して教授資格取得論文を執筆中とのことである。

本書は、3人の編者による序文で始まる。序文では、(1) 注意深く構築された大規模コーパスがあるのに、なぜ Web を使う必要があるのか、(2) なぜ代表性を備えていない Web を使うという危険を犯す必要があるのか、という問題に対して、以下の5つの点が挙げられている。

- ・ BNC (1億語) のような大規模コーパスでも、研究テーマ次第では十分な規模とは言えない。
- ・ ICE を除いて、既存のコーパスは英米の英語を中心に構築されており、世界各地の様々な変種が含まれているわけではない。
- ・ 技術的進歩によって新しいテキスト・タイプが誕生した(チャット、ブログなど)が、既存のコーパスには含まれていない。
- ・ 綿密に設計されたコーパスを構築するには、長い時間と多くの資金が必要とされるが、現在進行中の言語変化などを捉えるには、すぐに時代遅れになってしまう。
- ・ Web 上の言語自体が、言語変化に影響を与える可能性がある (weblish, netspeak)。

ここで述べられている点は、どれも説得力がある。現代英語の代表的コーパスである BNC は1億語であり、「大規模」コーパスと見なされていることは確かだが、1億語といっても書籍に換算して1,000冊程度である。大規模と言ってもこの程度である点を考えると、研究テーマ次第では Web に頼らざるを得ない場合があることは、当然と言えば当然である。ここに本書のような書籍が必要とされる理由がある。

序章に続く本論は4部構成になっている。第1部 (Accessing the web as corpus) は、3論文から構成され、Web をコーパスとして利用することの諸問題を扱っている。Google などの検索エンジンを使う長所と短所、WebCorp や KWICfinder などが紹介されている。

第2部 (Compiling corpora from the internet) の3論文は、インターネット上の資料を用いて自らの研究に必要なコーパスを構築する例が紹介されている。例えば、Hoffmann 論文は、CNN のトランスクリプトをダウンロードし、複数の Relational Database に格納しておけば、柔軟な検索ができることを示している。具体例として、強意の so が取り上げられている。中でも興味深く感じられた用法は、so not true 型である。これは、Hoffmann 自身の言葉によると、Quirk et al. (1985) には言及がなく Huddleston & Pullum (2002) が注において「比較的新しく、若い世代のインフォーマルな言葉遣いに特徴的に現れる」と述べている用法である。この種の用法は叙述形容詞の前で使われるのが専らであったが、最近ではそれ以外の環境でも使われるようになってきているとのことである。この用法が増加傾向にあることを示すグラフが載っているが、正に現在進行中の言語変化を捉えていて興味深い。

第3部 (Critical voices) は、Geoffrey Leech と Graeme Kennedy による2論文から構成されている。Leech 論文は、注意深く構築されたコーパスがもつべき要件(代表性など)が Web に対してどのように当てはまるか、あるいは当てはまらないかを検討している。(Google などの検索エンジンの中がブラックボックスである点にも言及されているが、これも重要な点である。また、検索結果が常に一定であるわけではない、という点も特に量的に利用する場合には留意しなくてはならない点である。) Kennedy 論文は、むしろ BNC が未だ十分に探求し尽くされていないことを述べ、Web に依拠する前にすべきことがある、というスタンスで書かれている。

第4部 (Language variation and change) には、7編の論文が収録されている。どれも Web を使わないことには、十分な資料が集まらない種類の現象が扱われている。紙面の都合で2編の論文にのみ触れる。Rosenbach 論文は、s 属格と NN 複合語 (lawyer's fees と lawyer fees) の相違を扱っている。既存のコーパス(但し、100万語か200万語のコーパス)を使っても十分な例が得られないことを示した上で、Google および WebCorp を使い、出現数を比較している。その上で所

有者の有生・無生が両者の区別に関わっていると主張する。また、Google は言語学者のための検索エンジンではないから、lawyer fees での検索結果の中に lawyer's fees が混在するなど、言語学者以外には重要でない区別に対しては無頓着である一方、WebCorp はその点で厳密であるから、言語研究においては後者の方が有益だという。しかし、WebCorp はサーチエンジンに依拠しており、その一部を利用しているのだが、どの一部を利用しているかが不明であるなどの問題点も併せて指摘されている。Anderwald 論文は、FRED コーパスと Google や WebCorp を用いたデータを用いつつ、動詞の周近的な過去形 (sink, drink などの過去形としての sunk, drunk 型) を扱っている。なお、第 4 部の諸論文はどれも興味深い現象を取り上げてはいるが、Google などの検索ヒット数を頻度と同一視している論文があるなど、少々問題含みのものもあることを指摘しなく

てはならない。

以上、本書の内容をごく簡単に見てきた。本書でも強調されているが、Web の場合には、研究において重要な再現可能性が確保できないという問題がある。Web を特に「量的に」利用しようとする場合、この問題は無視できない。しかし、この限界を逆手に取ることで、Web だからこそできる研究が存在することも事実だろう。それを行うにあたり、本書は参考になる。

なお、本書の出版年は 2007 年であるが、シンポジウムの開催年が 2004 年である点は要注意である。ここで述べられていることの全てが現在でもそのまま当てはまるとは考えにくい。むしろ、現在でも通用する見解ともはや通用しなくなった見解とを注意深く識別しながら読むことが大切になってくる。そうすることによってこそ、本書から様々な有益な示唆が得られると思われる。



英語コーパス学会 Newsletter No. 69

June 1, 2010

■会長:赤野 一郎
■事務局:〒739-8521 東広島市鏡山1-7-1 広島大学大学院総合科学研究科 井上永幸研究室内
■TEL:082-424-6431 ■振替口座:00940-5-250586(英語コーパス学会)
■URL: <http://muse.doshisha.ac.jp/JAECS/index.html> ■email: inoue@v.email.ne.jp

JAECS
Japan Association for English Corpus Studies

第35回大会報告

■概要

英語コーパス学会第35回大会は、4月24日(土)、兵庫県立大学(神戸学園都市キャンパス)で開催されました。快晴の中、神戸市営地下鉄学園都市駅下車徒歩約15分という立地条件の良さもあって、130名の参加がありました〔会員109名+新入会員2名+当日会員19名〕。

午前中のワークショップは「Web コーパス検索ツール Sketch Engine の基本操作と活用」と題して投野由紀夫先生(東京外国語大学)に講師を務めていただきました。Sketch Engine は、既成のコーパスを自由に利用できる Preloaded corpora (19言語26種類のコーパス群〔2010年4月11日現在〕); 英語は7種類のコーパスを含む)と自分で作ったコーパスを活用できる My corpora から成るシステムで、現在は Bank of English などでも利用されているインターフェイスです。本ワークショップでは、まず前半の基礎編で、アカウントの登録方法、基本的な検索方法、統計情報の取得方法などを扱い、後半の応用編では、CQL を使ったより高度な検索方法やコンコーダンスの用例を英文としてよりわかりやすい順に並べ替える方法の他、ワードリスト、ワードスケッチ、シソーラス、Sketch-Diff、Web Corpus などの各機能の作成方法について網羅的に解説いただきました。現在注目を集めている Sketch Engine というテーマに加えて、常にこういった最前線に身をおかれている投野由紀夫先生に登壇いただけるとあって、定員50名のところ、急ぎよ会場校のご配慮でブースを増加していただき、参加者は67名〔うち7名は補助椅子及び立ち見〕を数え大盛況となりました。参加者の皆様からも、非常に有意義なワークショップであったとの評価をいただいております。周到な準備の上で講師を務めていただきました投野由紀夫先生、そしてアシスタントを務めていただきました東京外国語大学の大学院生のみなさんには、この場をお借りしてお礼申し上げます。

午後の大会では、まず赤野一郎会長(京都外国語大

学)の開会の挨拶があり、その後、開催校を代表して濱田道夫先生(兵庫県立大学経済学部長)にご挨拶をいただきました。引き続き、加野まきみ先生(京都産業大学)の司会により年次総会が開かれ、2009年度の決算報告と会計監査報告、2010年度の予算説明があり承認されました。大会に出席されなかった会員の皆様には決算書と予算書を同封いたしますので、ご確認ください。最後に学会賞に関して、学会賞には2件の応募があり、奨励賞には応募が無かった旨、投野由紀夫学会賞選考委員長から報告がありました。引き続き「言語研究」と「英語教育」のテーマで2室に分かれて、五百蔵高浩先生(高知女子大学)と阿部真理子先生(高崎経済大学)の司会のもと、それぞれ2件ずつの研究発表が行われました。今大会では、その後、山崎俊次先生(大東文化大学)の司会で、大名力先生(名古屋大学)によるご講演「コーパス検索で注意すべきことー基礎データの信頼性向上のためにー」が行われました。最後に、午前中のワークショップと連動して、投野由紀夫先生の司会のもと、「進化する Web コーパス:現状と課題」と題するテーマでシンポジウムが行われました。それぞれの司会の先生にご執筆願いました概要につきましては、「研究発表」、「講演」及び「シンポジウム」のセクションをご覧ください。最後に、開催校の瀬良晴子先生(兵庫県立大学)の開会の辞をもって、第35回大会の幕を閉じました。

大会終了後の懇親会には54名の参加がありました。家口美智子先生(摂南大学)の司会のもと、会長挨拶の後、前会長の中村純作先生(立命館大学)の乾杯のご発声で懇親会が始まりました。会員同士の交流と情報交換で盛り上がり、午後8時半にはすべての大会行事が終了いたしました。

最後に、本学会員で開催校責任者の瀬良晴子先生のご尽力と細部にまで配慮の行き届いたご協力で盛会に終わったことを喜び、心よりお礼申し上げます。加えて大会実施に協力くださった学生、院生の方々にもこの紙上を借りて厚くお礼申し上げます。

■研究発表

*no doubt*の機能について：類義表現との比較から

鈴木大介（京都大学大学院/日本学術振興会）

本発表では、英語の *no doubt* という表現の語用論的特性を明らかにする試みが報告された。Simon-Vandenberg and Aijmer (2007), Biber et al. (1999) といった先行研究では、蓋然性や確信度といった質的分析が多くなされていること、談話標識としての *no doubt* も存在するとあるように、*no doubt* に関わる意味の広がりが見られていること、他の副詞との違い、すなわち *no doubt* 特有の働きについては示されておらず、量的な研究が不十分で、使用状況についての全体像は未だ不明確なままであることがまず指摘された。その上で、BNC (BYU-BNC) を用いた *no doubt* の使用パターンを明らかにする試みが報告された。

コーパスから *no doubt*, *doubtless*, *undoubtedly* について得られた用例数は、それぞれ、5,955 例、844 例、2,343 例であった。文副詞として生起しているものを手作業で調べ、各々2,701 例、734 例、2,202 例を抽出した。これをサンプルとし、「談話構造」「モダリティ」という2つの側面から分析が行われた。

「談話構造」の観点から前後のメッセージのつながりの関係を明確に表す指標に着目し、(i) *no doubt* と *doubtless* が節中のどの位置に生起しているか、(ii) それぞれの節の主語が代名詞であるかどうか、(iii) それぞれの節頭に談話標識が生起しているかどうかを変数とした調査結果と、「モダリティ」の観点から、話者の心的態度を明確に表す指標に着目し、両表現が生起する節内において、(i) 法助動詞、(ii) 法副詞、(iii) comment clauses (*I'm sure, I suppose, it seems, ...*) との共起頻度の調査結果が報告された。

結果、*doubtless* は節中に生起している割合が高い一方で、*no doubt* は、節頭の割合が非常に高く、談話機能を担いやすいことが示された。*doubtless* と比べ、*no doubt* の節において代名詞が主語となる割合が高いことが提示され、*no doubt* が談話の流れの中で生起しやすいことが示唆された。談話標識との共起については *no doubt* が他の2語と比べ高い共起頻度を示した。しかし、統計的有意差は見られなかった。心的態度を表す表現との共起については *doubtless* の高い共起頻度が確認された。意味領域では、*certainly*, *probably* を表し、*doubtless*, *undoubtedly* と類義である。しかし、節を超えた談話語用論レベルにおいては、担う機能の種類に差異があることが指摘された。*doubtless* はモーダル機能を担うのに対して、*no doubt* は談話機能を担う傾向が強いと考えられることが示された。

最後に、*certainty* と *probability* の区別、*no wonder* 等の他の「no+名詞」の形式との関係、節末の *doubt* と

spoken の関係、*no doubt* の持つ意味・機能について、配付資料の用例の解釈などについてフロアから質問および指摘があった。

五百蔵高浩（高知女子大学）

古英語・中英語に現れる小節・結果構文：通時的英語コーパス YCOE と PPCME 2 を検索して

神谷昌明（豊田高専）

本発表では古英語、中英語に同じコーパス デザインで文法標識が付与された2つのコーパス、The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE; 古英語) と The Penn-Helsinki Parsed Corpus of Middle English Second Edition (PPCME2; 中英語) から結果構文を抽出し、抽出した用例を先行研究の記述に追加する試みが報告された。結果構文は歴史的に見れば古英語 (OE) の動詞接頭辞構文 (小辞構文) や中英語 (ME) 期に生産的になった句動詞 (phrasal verb) 構文と大いに関連があると言われているが、通時的に英語の結果構文を詳細に分析した先行研究は少なく、共時的研究が多数を占めていることが指摘された。本研究では、結果構文の史的発達 (成立過程) を探る通時的研究へ繋げるための準備段階として、古英語および中英語に現れる結果構文の具体例を抽出する試みが行われた。結果構文を含む小節を示す統語標識 [IP-SMC] を検索文字列としたコンコーダンス作成と手作業による該当構文の抽出が説明された。結果、19 作品から、比喩的表現などを含めた32例の結果構文が得られ、中英語期の結果構文について Visser (1967) が挙げる17動詞に加えて、さらに12動詞 (*baken, striken, schitten* 等) を新たに見つけ出すことができたことが報告された。しかし、*He broke it into pieces* のような結果構文を検索することが現状では困難であり、数ある小辞構文の中から結果構文のみを抽出する方法をさらに検討していくことが課題であると報告された。発表後、フロアからは次のような質問および指摘があった。(i) 小節のタグ [IP-SMC] では結果述語になりうる前置詞句が検索できないのなら、動詞のコンコーダンスを作成しその後継続前置詞句を調べれば探し出せるのではないか。(ii) 古英語では従属節の中で、語順の変化が起こりうるので、結果述語が動詞の前に来るのはさほど稀な現象ではないのではないか。(iii) 結果述語は認知の観点からは目的語の後に来た方が理にかなっているように思えるが、古英語・中英語では挙げられている用例が示すように様々な位置に結果述語が現れている。

五百蔵高浩（高知女子大学）

コーパスを応用した言語テスト問題の改良 —日本の大学入試問題から文法4択問題を例に一

宇佐美裕子 (ランカスター大学大学院生)

コーパスをもとにした語彙・文法の記述は、言語テストの分野にも応用されてきているが、一般コーパスや学習者コーパスがテスト作成や採点などに十分に利用されているとは言い難い。しかし最近ではコーパスを言語テストへ応用した研究も行われるようになってきた。その一例として、Educational Testing Service (ETS) や Cambridge ESOL は、独自のコーパスをもとにテストを作成し、評価を行っている (e.g. Biber et al. 2004, Barker 2005)。この発表では、日本の大学入試問題が信頼性や妥当性の点で問題のあるアイテムを含んでいる (e.g. Brown 1996, Watkins et al. 1997) という点に着目し、コーパスを利用することで、より良いテスト作成ができるのではないかと提案がなされた。

コーパスとしては、研究発表者によって構築された 2002 年から 2007 年に出題された私立大学入試問題とセンター試験問題からなる 133,000 語規模のデータが利用された。そして文法空所補充 4 択問題の中でも、比較的多く出題され、日本人学習者のエラーが多く見られたイディオム *in spite of* (誤用例: *In spite of I bullied her, she helped me.*) に関して分析が行われた。まず、*in spite of* が大学入試問題の stem に使用されている例 “*In spite of her illness, ...*” と、British National Corpus に現れている例 “*In spite of the fact that ...*” の違いが指摘された。そして BNC に出現しているコロケーションを利用した新しい stem (例: *Do you love me in spite of the fact that I haven't succeeded?*) が提案された。さらには、大学入試問題に使用されている錯乱肢 (例: *In spite of / Instead of / Thanks to / Because of*) と、Longman Learner Corpus に観察された日本人学習者の *in spite of* に関するエラーを比較し、学習者のエラー傾向を考慮に入れた新しい錯乱肢 (例: *In spite of his sickness / Despite of he is sick / In spite of his sickness / In spite of he is sick*) の提案がなされた。

質疑応答では、何を基準に良いテストとみなすのかという点とコーパスデータをテスト作成に利用することで、信頼性や妥当性の問題が解消されるのかという点が今後の課題となるのではないかと指摘があった。また入試問題では、*in spite of* の使い方を問題にしているのではなく、意味が正確に理解できているかを試しているのであり、何が問題ごとに問われているのかを分析する必要があるのではという指摘があった。今後、日本においてデータ収集が行われるということであったので、さらなる研究成果が期待される。

阿部真理子 (高崎経済大学)

「日本語母語英語学習者の連語表現の判別分析」

石田知美 (名古屋大学)

杉浦正利 (名古屋大学)

英語母語話者と英語学習者の連語表現の使用にどのような特徴があるかを明らかにすることが目的とされていた。NICE (Nagoya Interlanguage Corpus of English) の英語学習者コーパスと母語話者統制コーパスの中でも、同じトピックで書かれた作文データが比較分析の対象とされた。まず 2 語以上の連語表現の使用について、(a) 総数、(b) 異なり数、(c) Guiraud Index (GI) の点から、学習者と母語話者の差異を明らかにした後、判別分析を用いて、両者の判別に寄与する項目が検証された。

結果として、以下の 2 点が挙げられた。(1) 学習者の方が総数・異なり数・GI のすべてにおいて有意に多い。(2) 学習者の方が長い連語表現を使用する。この傾向については、学習者の用いる連語表現のレパートリーが少なく、同じ表現が繰り返し使われるためではないかとの指摘があった。また学習者の方が異なり数・GI が多いという点については、一度しか使用されない連語表現をカウントしなかったため、語彙の多様性を表す指標である GI をそのまま使用しても、表現の多様性をはかることができないためではないかとの指摘があった。

さらに、学習者の方が長い連語表現を使用するという結果を受けて、6 語表現を例にとり、具体的にどのような表現が使用されているのかが分析された。結果、英語母語話者は 14 エッセイ (全体の 18% 分散率 0.74) においてしか、6 語表現を用いていないのに対し、学習者は 51 エッセイ (全体の 41% 分散率 0.84) において、6 語表現を用いていることが明らかになった。しかしながら、学習者が使用している 6 語表現は、文法的には非文ではないものの、英語母語話者には修正ポイントとしてあげられているケースが多いことも分かった。

判別分析の結果、2 語表現は「総数・異なり数・GI」の 3 変数とも、学習者と母語話者の判別に寄与したが、3 語表現は「総数・異なり数」のみ、4 語表現と 5 語表現は「総数」のみ、そして 6 語表現から 8 語表現までは「GI」のみが、判別に寄与することが明らかになった。そして、語レベルで判別分析を行った Sugiura et al. (2007) と合わせて考察すると、5 語以下の連語表現であれば、連語表現と単語の総数を観察することによって、学習者と母語話者の判別が可能になることが示唆された。

質疑応答では、統計処理に関する質問から、母語話者の定義に関する質問まで、幅広い問題提起があった。今後の課題としては、表現の多様性を示す新たな指標が必要となることが明らかになり、学習者と母語話者の判別に寄与する言語的特徴に関するさらなる研究成果が期待される。

■招待講演

コーパス検索で注意すべきこと —基礎データの信頼性向上のために—

大名 力 (名古屋大学)

本講演は、コーパスを利用した研究では、大規模コーパスをコンピュータ処理し、さらに統計処理も行われるため、客観的で信頼性が高いという思い込みから、批判的な検討なしにコーパスを利用し、先行研究のデータ、結論を無批判に受け入れてしまうことの危険性を指摘したものであった。そうした状況を改善するためには、批判的な検討を行っていくとともに、研究者間での情報共有をより一層進めていくことが重要であることを訴えたものであった。

講演の前半では、特別な知識や技術がなくてもコーパスが利用できるユーザーフレンドリーなツールの普及はコーパス研究の向上に寄与してきた反面、コーパス、処理内容、手法・方法論のブラックボックス化を進める要因の1つにもなってきたことを強く指摘した。さらに、研究目的に合った正しいデータを得るには、コーパスやテキスト処理、統計処理に関する知識・技術は不可欠であると力説した。

講演の後半では、コーパス研究の基礎データとなる出現頻度に焦点を当て、コーパス研究者が無意識のうちに犯しやすいミスについて具体例をあげて解説した。例えば、サーチエンジンを使う場合、出現数の信頼性や解釈の問題について、Brown Family of Corporaでは、2行にまたがるコロケーションの処理、行頭のテキストID等の処理、コーディングの問題等を提示し、また、British National Corpusでは、ヘッダー部分の処理、phantom words (e.g. “<w VM0>wo<w XX0>n't”, cf. “<w VM0>will <w XX0>not”), multi-words (e.g. “<w PRP>but for” vs. “<w CJC>but <w PRP>for”)の処理等の問題点を指摘した。さらに、研究書、言語学書等からの収録テキストに含まれる作例・非文の扱いの問題等を取り上げ、コーパス検索やデータの分類・解釈において注意すべき点について具体的に説明した。

最後に、まとめとして、研究者が注意すべき点を網羅的に列挙するのは不可能であり、個々の研究では研究者自らが注意深く考えて適宜対応する必要があることを述べた。その為には、①作業内容の明確化(特に、入力・処理・出力をセットで考えること)、②コーパス、ソフトのマニュアルへの精通、③コーパスおよび抽出データの直接読みの励行が重要な課題と説き、多様な規模・種類のコーパスとユーザーフレンドリーなソフトが容易に入手可能になった現在だからこそ、批判的な検討が必要であり、研究者間で情報を共有し

ていくことが重要であることを強調し、講演を終了した。

山崎俊次 (大東文化大学)

■シンポジウム

シンポジウム《進化する Web コーパス：現状と課題》

Web コーパス概観

司会 投野由紀夫 (東京外国語大学)

コーパス言語学・自然言語処理の分野で近年注目を集めている Web コーパスを特集するシンポジウムが開催されるのは意義深いことである。ここではイントロダクションとして、まず Web コーパス発達の歴史的な概観、現在の Web コーパス構築の一般的な手法の解説、利用できる Web の規模とコーパスの規模の拡大の様相などをまとめた。その後、コーパス言語学の目的に照らした場合に、Web コーパスがコーパスとしての条件(代表性や均衡など)を的確に満たしているのか、研究対象の言語事象とコーパスの最適規模の問題、巨大データをいかに高速に検索するか、といった課題を提示し、講師の先生方への序章とした。

Web コーパスの言語情報処理基盤

講師 田中省作 (立命館大学)

近年、膨大な Web 情報の効率的収集・高度情報処理に向けた共通研究基盤や各種要素技術の開発が進められている。これらが今後の Web コーパス研究において重要な役割を果たすことは間違いない。そこで本発表では、このような国内プロジェクトの現状を概観し、個々人では対応が難しい検索エンジンの不透明性といった諸問題の解消が期待されることを述べた。また、Web コーパス研究の事例として文章の質情報を付した英語科学論文コーパスを挙げ、個別研究における目的に応じたより細かな情報処理技術の必要性を論じた。

Web コーパス研究におけるブログの可能性とその文体的特徴

講師 大羽 良 (中央大学)

本発表ではブログからコーパスを作成し、その文体的特徴を分析した。具体的にはウェブサービス“technorati”のトピックを言語外要因とみなし、言語内要因のひとつである単語の生起頻度との関係に対応分析を用いて分析を試みた。結果として以下の点が明らかになった。(1) ブログコーパスのジャンルの類型化はBNCのテキストタイプの類型化に近似した。(2) BNCを含めた類型化ではBNCの書き言葉が分布の中心に位置し、ブログコーパス特有の頻度の偏りはないと推

定される。(3) 大量のデータから成るブログコーパスを用いても、Brown・Frown コーパスを適切に分類することはできなかった。これは大量のデータを用いても、適切なジャンルを収集しなければジャンルを分類することは不可能であることを示唆する。

小学館 Web コーパス “Sekai Corpus” 構築とその活用

講師 中村隆宏 (ネットアドバンス SE)

星野 守 (小学館コミュニケーション編集局)

小学館では社内研究開発用途に Web から多数の専門分野を集めた Sekai Corpus を開発した。本発表の前半では、バランスコーパスを目的にしたランダムなキーワードによるランダムサンプリングの技術的な難しさについて述べ、その解決方法として、標準的な語彙順位リストに近似した語彙頻度分布となるようにサンプリングするという提案を行った。後半では、Sekai Corpus のサブコーパス比較について概説した。Sekai Corpus では、100 万語あたりの使用頻度の多寡で KWIC 行数を調整し、サブコーパス同士を直接に比較できるようにしている。また、共起調査用の語句集計においてもサブコーパスの分布が直観的に把握できるようなインターフェイスにし、観察者の理解を助ける工夫をしている。今後は、収集分野をさらに充実させるとともに、上記で示した手法による標準コーパスの開発を進めて Sekai Corpus に組み入れる予定である。

新入会員紹介 (5 月 17 日現在, S は学生)

柴崎礼士郎	沖縄国際大学
福田 稔	宮崎公立大学
古樋直己	大阪大学大学院 S
堀田隆一	中央大学
宮澤信一郎	秀明大学
山下美朋	立命館大学大学院 S

第 36 回大会研究発表者募集

2010 年度秋季大会 (第 36 回大会) は 10 月 9 日 (土) に東京大学 (駒場キャンパス) で行われる運びとなりました。つきましては、発表を希望される方は、下記の要領に従って email で事務局宛にお申し込み下さい。

【分野】 本学会にふさわしい、コーパス利用・コンピュータ利用を中心に据えた英語研究。

【応募資格】 本学会員であること。

【発表方法】 発表 20 分、質疑 10 分。

【応募方法】 冒頭に題名のみを記し、800-1200 字 (参考文献は別) にまとめ、メール添付ファイルで送付。メール本文に氏名 (ふりがな)、所属・職名、住所、電話番号、メールアドレス明記。

※審査の際、応募者が特定されないよう、事務局が応募書類を加工させていただくことがございます。

【応募締め切り】 2010 年 6 月 15 日 (火) 必着

【採否決定】 2010 年 7 月下旬 (予定)

【問合せ】 〒739-8521 広島市鏡山 1-7-1

広島大学大学院総合科学研究科

井上永幸研究室内

英語コーパス学会事務局

email: inouc@v.email.ne.jp

※問合せ先が変更になっております。ご注意ください。

ハンドアウトのダウンロードサービス

第 35 回大会の研究発表とシンポジウムのハンドアウトを希望される会員に対して、ダウンロードのサービスを行います。期間は、このニューズレターお届けより 6 月 30 日までとします。ファイルは PDF となっております。ご希望の方は、石川保茂 (yasuishikawa@hotmail.com) まで下記のハンドアウトのうちご希望の番号をお知らせください。追って URL をお知らせいたします。

なお、発表者の著作権保護の立場から印刷は「許可しない」に設定してあります。

※以下、講師・発表者の敬称は略させていただきます。

- 1 投野由紀夫「Web コーパス検索ツール Sketch Engine の基本操作と活用」
- 2 鈴木大介「no doubt の機能について」
- 3 神谷昌明「古英語・中英語に現れる小節・結果構文」
- 4 宇佐美裕子「コーパスを応用した言語テスト問題の改良」
- 5 石田知美・杉浦正利「日本語母語英語学習者の連語表現の判別分析」
- 6 大名力「コーパス検索で注意すべきこと」
- 7 投野由紀夫「Web コーパス概観」
- 8 田中省作「Web コーパスの言語情報処理基盤」
- 9 大羽良「Web コーパス研究におけるブログの可能性とその文体的特徴」
- 10 中村隆宏・星野守「小学館 Web コーパス "Sekai Corpus" 構築とその活用」

会誌『英語コーパス研究』第 18 号論文投稿募集について

本ニューズレターとともに、『英語コーパス研究』第 17 号が同封されています。第 17 号には 7 点の論文と 1 点の研究ノートが採択されました。

続きまして、『英語コーパス研究』第 18 号の原稿を

次の要領で募集いたします。会員各位の積極的な投稿をお待ちしております。

【原稿の種類】

1. 英語コーパス利用・コンピュータ利用を中心に据えた「研究論文」、「研究ノート」、「実践報告」。
2. 「書評」、「コーパス紹介」、「ソフト紹介」、「海外レポート」、「論文紹介」などの各種情報あるいは紹介原稿。

【投稿申込締切】2010年7月31日（土）

※氏名、所属、原稿の種類とタイトルを下記原稿提出先まで電子メールにてお知らせください。

【原稿提出締切】2010年9月30日（木）

※ハードコピー4部を郵送。およびWORDファイルを電子メール添付にて提出（電子メール件名には「英語コーパス研究 18号投稿論文」とし、メール本体には、氏名、所属、原稿の種類とタイトルを書いて下さい）。論文、研究ノートの冒頭には題名のみ記し、氏名（ふりがな）、所属・職名、住所、電話番号、電子メールのアドレスを明記した別紙（1部）を添付のこと（WORDファイルもこの形式に沿ってください）。

【問い合わせ先・原稿提出先】

〒980-8576 仙台市青葉区川内 41
東北大学・大学院国際文化研究科 岡田 毅
TEL: 022-795-7632 FAX: 022-795-7632
Email: t-okada@intcul.tohoku.ac.jp

【原稿の長さ】

1. 研究論文
和文 A4 サイズ 1 ページあたり 35 字×30 行、17 枚以内。
英文 A4 サイズ 1 ページあたり 70 ストローク×35 行、17 枚以内（10.5 ポイント使用）。
※いずれも Abstract（英文）、図表、注、書誌、付録を含む。
2. 研究ノート
1 ページあたりは上記の書式と同様で、12 枚以内。
※いずれも Abstract（英文）、図表、注、書誌、付録を含む。
3. その他
研究論文の半分以下。

【書式】

第 17 号所収の論文を参考にしてください。詳細は学会ホームページ（<http://muse.doshisha.ac.jp/JAECS/>）でご確認ください。

【採用通知】2010年11月頃

【刊行予定】2011年5月

なお、7 月末に設けられた投稿申込締め切りへの募

集の有無に関わらず、9 月末の原稿締め切りまでに投稿頂ければ、会誌への投稿は可能です。

『英語コーパス研究』編集委員会委員長
岡田 毅（東北大学）

人事に関する決定事項について

大会前日の 4 月 23 日午後 5 時 30 分より開かれた運営委員会において人事案が審議されました。まず、山崎俊次先生（大東文化大学）が規定（一期 2 年）を鑑み、2010 年度も副会長を継続されることとなりました。

次に、運営委員の継続に関してですが、昨年度で任期の切れる 5 名の先生方のうち 4 名〔新井洋一先生、岡田毅先生、高橋薫先生、塚本聡先生〕の留任、1 名〔大津智彦先生〕の退任が承認されました。お疲れさまでした。

編集委員としては、岡田毅先生（東北大学）、滝沢直宏先生（名古屋大学）、小林多佳子先生（昭和女子大学）の 3 名の先生方が再任され、園田勝英先生（北海道大学）が新しく承認されました。また、新しく編集委員長として岡田毅先生の就任が承認されました。2 期 4 年間にわたって編集委員長を務めてこられた塚本聡先生（日本大学）には、ご尽力に対し心よりお礼申し上げます。

学会賞選考委員としては、長年担当された金子朝子先生（昭和女子大学）、西納春雄先生（同志社大学）の後任として、中條清美先生（日本大学）、成田真澄先生（東京国際大学）が承認されました。

さらに、大会企画委員会の新設により、滝沢直宏先生（名古屋大学）、石川慎一郎先生（神戸大学）、後藤一章先生（摂南大学）、阿部真理子先生（高崎経済大学）、金沢俊吾先生（高知県立高知女子大学）、西村秀夫先生（姫路獨協大学）の 7 名の先生方が委員として承認され、滝沢先生に委員長を務めていただくこととなりました。

東支部の支部長としては、投野由紀夫先生（東京外国語大学）が留任されることとなりました。

年 1 回の研究大会への移行

既に述べました 4 月 23 日の運営委員会におきまして、これまで春と秋の年 2 回開催して参りました研究大会を、より充実したものにすべく、2011 年度より年 1 回 2 日間にわたる開催へと移行することが認められました。今後の発展にご期待下さい。

学会賞について

3 月 31 日に締め切りました今年度の学会賞候補には 2 件の推薦がありました。また、奨励賞には応募があ

りませんでした。今年度秋の大会で結果を発表すべく、現在選考委員会で鋭意審査中です。

学会賞選考委員会委員長
投野由紀夫（東京外国語大学）

東支部活動報告

東支部では昨年度、ワークショップを1回と特別講演を1回開催しました。2009年11月8日（日）に大東文化大学を会場に、開発者の Adam Kilgariff 氏をお招きして Sketch Engine のツール講習会を約30名ほどの参加者を得て開催。皆さん、熱心に学んでいました。さらに、12月8日（火）には同じく大東文化大学で、Douglas Biber 教授の特別講演会を開催。大東文化大学の学生、英語コーパス学会の会員など総勢80名ほどが熱心に聴講しました。会場のお世話をいただいた山崎俊次先生と鈴木敬了先生にこの場を借りて御礼申し上げます。

今年度は、ツール講習会の継続、講演会や読書会などの企画を運営委員の先生方と相談しながら、できる範囲で企画していければと思っております。また、何かアイデアなどあれば是非お聞かせください。よろしくお願ひします。

東支部支部長
投野由紀夫（東京外国語大学）

今後の大会日程と開催校

第36回大会	2010年10月9日（土）
東京大学（駒場キャンパス）	
第37回大会	2011年10月
開催校未定	

事務局から

◇会則の改正

会則が一部改正になりました。新しいものを同封しておりますのでご覧ください。

◇会費納入のお願い

2010年度会費（一般5,000円、学生3,000円）を、日本郵便にある払込取扱票を使ってお納めいただきますよう、ご協力をお願いいたします〔振替口座：00940-5-250586〕。日本郵便発行の受領証をもって領収書に代えさせていただきますので、ご了承ください。別途領収書が必要な方は、80円切手を同封の上、石川保茂（〒615-8558 京都市右京区西院笠目町6 京都外国語大学）までお申し出ください。払込取扱票の通信欄によるお申し出はご遠慮ください。

過年度会費未納の方は、2010年度分と併せてお納めください。会誌『英語コーパス研究』第17号は2009

年度の会費を納入していただいた方のみ、送付いたします。また、2年続けて会費未納の場合、Newsletterなどの送付を中止させていただきます。

住所、所属などに変更や異動のある方は、必ず払込取扱票の通信欄にお書き添えください。

※会員の皆様には、日頃より会費の当該年度内納入のご協力をいただきまして、お礼申し上げます。会費を滞納されますと、退会時に滞納分をまとめてお支払いいただくといった事態にもなりかねません。会員の皆様におかれましては、円滑な学会運営のためにご協力いただけましたら幸いです。

◇メーリングリストについて

英語コーパス学会ではメーリングリストを使って会員の皆様の様々な情報交換に役立てていただいているところですが、最近、宛先不明でエラーが返ってくる例も増えています。会員の皆様方には、メールアドレスに変更が生じた場合、速やかに事務局宛ご連絡いただけますようお願い申し上げます。

☆☆☆☆☆☆☆☆☆☆

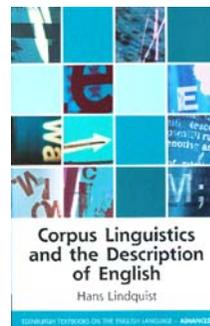
FORUM

■文献紹介

赤野一郎（京都外国語大学）
i_akano@kufs.ac.jp

Lindquist, Hans (2009) *Corpus Linguistics and the Description of English*. Edinburgh Textbooks on the English Language – Advanced. Edinburgh: Edinburgh University Press. 224 pages.

£18.99 (paperback)
ISBN-10: 0748626158
ISBN-13: 978-0748626151



本書はエジンバラ大学出版局から Edinburgh Textbooks on the English Language – Advanced の第5巻目として出版された英語コーパス言語学の入門書である。著者の Hans Lindquist はスウェーデンの Linnaeus 大学の准教授で、Lindquist & Mair (2004) の共編者である。前書きの To readers によれば、コーパス言語学との最初の出会いは、1977年 Lund 大学の大学院生時代、Jan

Svartvik の指導の下に London-Lund Corpus のタグ付け作業に関わった時とのことである。

本書の章構成は以下のようになっている。

1. Corpus Linguistics
2. Counting, calculating and annotating
3. Looking for lexis
4. Checking collocations and colligations
5. Finding phrases
6. Metaphor and metonymy
7. Grammar
8. Male and Female
9. Language change
10. Corpus linguistics in cyberspace

最初の 2 章は後に続く 8 章への導入の役割を果たしている。残りの 8 章のうち、3 章～5 章はコーパス言語学の中心的な研究課題の 1 つである語の振る舞いに関する内容に当てられており、6 章～9 章は語を超えた言語単位の分析に当てられ、‘Web as Corpus’ を扱う 10 章で締めくくられている。

各章は本文、Study questions、Corpus exercises、Further reading の 4 つの部分からなっている。Corpus exercises では、それぞれの章の内容に合わせた hands-on exercise が出版社のホームページ (<http://www.eupublishing.com/series/ETOTELAdvanced/Lindquist/>) 上に用意されており、このサイトの指示に従って練習問題をこなすだけでも、種々のコーパス (BYU-BNC, COCA, TIME, MICASE, WebCorp) の利用法や Google の活用法が体得できる。以下、紙幅の許す範囲で各章を概観する。

第 1 章では Brown Corpus に至る実証主義的言語学の流れが概観され、コンコーダンスの機能の説明とコーパスに対する初期の批判を踏まえ、Svartvik (1992) に基づきコーパス言語学の利点が提示される。章の後半ではコーパスの分類と利用可能な各種コーパスの特徴が手際よくまとめられている。

第 2 章は量的研究の基本である頻度について、BNC から得られた語の頻度表を示しながら、カイ二乗検定、サブコーパス間の分布、頻度の標準化、代表性について解説される。章の後半では、アノテーションについて POS tagging を中心にその有効性が示され、タグ自動付与プログラム、CLAWS tagger の必要性が述べられる。頻度と統計分析の重要性を認めつつも、著者は “In order to make the best use of the figures, it is vital to go back to individual examples from the corpus to look at the reality behind the figures.” と述べ、質的研究の重要性を強調している。

第 3 章では、まず辞書編纂におけるコーパス利用の現状を紹介し、LDOCE1 (1978) の特定の語彙項目の記

述とコーパスの検索結果を比べることで、辞書編纂者の経験や直観と、コーパスから得られた実際の言語使用の間にずれがあることを例証している (ただしコーパスに依拠した LDOCE3 (1995) と比較していれば、その結果は異なったものになっていただろう)。次に語が別の語と連語することで全体として感情的色合いを帯びる現象 semantic prosody (例: COCA における *set in*)、語の使用頻度の時代的変遷 (例: TIME Corpus に見られる *greenhouse effect* と *global warming*, COCA の subcorpora 間の *maybe* と *perhaps*)、各種新聞の CD-ROM を利用した変種間の使用分布調査 (例: *fatigue* を主要語とした複合語) が示され、最後に文体研究へのコーパス利用についての言及があり、特定の作家の特徴語を抽出する手法や登場人物間の特徴語の差異をあぶり出す事例が紹介されている。

第 4 章ではコロケーションを扱う。コロケーションに、Firth (1957) のある語と隣接する語 (collocates) との関係と、Palmer (1933) の「1 つの単位として学習すべき 2 語以上の連続」の 2 つを認め、前者を ‘collocations in a window’、後者を ‘adjacent collocations’ と呼び、別々に扱っている。前者は中心語 (node) の前後 4-5 語の幅 (span) において、結びつきの強さを測る統計指標 (MI, z-score など) を用いて共起語を抽出する方法で、*dog*, *educational* の分析事例が示されている。‘adjacent collocations’ は、一定の幅に現れる語を対象とするのではなく、その名が示すように、ある語の直前あるいは直後の語との結合傾向を見る。*bed* を含む 3 語連鎖、*hand* を含む 2 語連鎖の分析例が示されている。中心語と文法範疇の関係、すなわち colligation についても簡単に触れられ、Phrase in English (PIE) を利用し、*deny* で始まる 3 語連鎖を抽出し、おもな文法的パターン (*deny that*, *deny NP*, *deny doing*, *deny NP1 NP2*) を提示している。

第 5 章はある程度固定した (fixed) 語の連鎖、句表現 (phraseology) の重要性を論じている。‘semi-preconstructed phrases’ が言語理解の中心であるという、Sinclair の the idiom principle の説明に始まり、固定度の最も高いイディオムの分析例に移る。イディオムは、構成要素の意味を合成してもイディオム全体の意味にならないという非合成性 (non-compositionality) と構成要素の一部を変更できないという不変性 (invariability) を備えていると言われるが、“Very few idioms are 100% fixed.” と主張し、合成的なイディオムとして *high and dry*, ヴァリエーションのあるイディオムとして *a storm in a teacup* をあげて例証する。次いで Anteborg (1998) から 3~4 語からなる語連鎖で頻度の高い句表現 (recurrent phrases) の談話における機能分析を示した後、Phrases in English database の概要が紹介さ

れている。このサイトでは BNC から高頻度の n-gram の表現, ‘the * of the’ などのフレームに合致する句, ‘ART ADJ NOUN’ のような品詞連鎖に合致する高頻度の句などを取り出すことができる。本章では, *at the end of the* のような 5-gram の句, *toe* を含む 5-gram, *good* + 名詞の例があげられている。最後に Dickens の小説で頻繁に現れる句表現と他の作家の句表現を比較し Dickens の文体的特徴を明らかにする事例研究として Mahlberg (2007) を紹介している。

第 6 章は *metaphor* と *metonymy* をコーパスでいかに分析するかが扱われている。*metaphor*, *simile*, *conceptual metaphor*, *metonymy* について解説した後, コーパスに基づくメタファーの抽出と分析の方法を示す。具体的方法としては, たとえば概念メタファー, THE MIND IS A MACHINE においてメタファーの基になる ‘source domain’ の *machine*, あるいはメタファーの対象となる ‘target domain’ の *mind* と関係があり, メタファーとして使われそうな語 (e.g. *rusty*) やパターン (e.g. *his/her/my mind was a/an . . .*) をコーパスで検索し, メタファー表現がどうかを確認していくという方法である。紹介されているもう一つの方法は, サイズの小さなコーパスの中から手作業でメタファーあるいはメタファータイプを探していく方法で, アメリカ歴代大統領の就任演説を source domain ごとに分類した Charteris-Black (2004) が事例研究としてあげられている。さらに著者はこの研究から得られたメタファータイプが大規模コーパスでも典型的タイプであるのかどうかを検証することで, 小規模コーパスによる手作業の分析方法が有効であると主張する。コーパスに基づくこの分野の研究は始まったばかりで, ここで紹介されている分析方法は, 場当たりの感が否めないが, コーパス利用の幅を広げるためにも, さらなる研究が望まれる。

第 7 章は文法を扱う。これまでの章は語が分析対象だったので, タグがなくても語形をたよりに検索することができたが, 統語構造が検索対象となれば, タグ付きコーパスが不可欠である。ただし, この章で最初に採り上げている, 形態論に属する *nbo* と *nbom* の時系列での増減調査は語形検索で可能である。統語的分析の事例としては, TIME Corpus による 1920~2000 年代に至る期間における *get-passive* の使用頻度の変化, *be accustomed to do/doing* に関する Rudanko (2006) の研究結果 [かつては *to do* が普通であったが 20 世紀には *to doing* が圧倒的になった] と, 著者が行った COCA と BNC による検証結果の比較, *have difficulty in doing* における *in* の脱落に関する詳細な分析が含まれている。最後に ICE-GB のような *parsed corpus* が構文検索に欠かせないことにも言及されている。

第 8 章ではコーパスを使って英語とジェンダーの間

題を探る試みが紹介されている。*firemen* のような男性を無標とする職業名詞が嫌われ, 中立的な *firefighter* に取って代わられてきていることは周知の事実だが, TIME Corpus を使ってこの種の名詞の使用の変化を示している。また形容詞に着目し, *man* と *woman* がどのように描かれているか, 動詞に着目し男性と女性がそれぞれ典型的に行う行為はなにかなどを分析している。ことばに見られる性差の例としては, 色彩語は女性のほうが繊細な使用が見られる, 垣根表現は女性が多用するとよく言われるが, コーパスを使って検証している。色の問題については BNC における男性の書き手と女性の書き手の間に見られる色彩語の分布を調査することで確認できたが, 後者に関しては *you know*, *kind of*, *sort of* の男女間の頻度差を調査したが, 断定的なことは言えないとしている。

第 9 章では英語の変化を採り上げている。著者は言英語の変種間の時間的変化 (例: 副詞としての単独使用の *likely*) や世代間の相違も含めている。通時的な研究方法の 1 つとして, 先行研究といくつかのコーパスを組み合わせることで, 全体の変化を明らかにする方法を述べている。Helsinki Corpus, ARCHER Corpus, FLOB, Frown, BNC に基づき *beside(s)* の文法化を論じた Rissanen (2004) を最初に紹介し, 1990 年あたりで終わっているこの研究を引き継ぐ形で, TIME Corpus を活用し 1920~2000 年代の変化を調査し変化の全体像を示している。もう 1 つの方法として, OED をコーパスとして利用する方法が, *start to do*, *start doing* に見られる *start* の準助動詞化を明らかにした Mair (2004) を引用しながら説明される。最後に社会言語学的分析事例として, Nevalainen & Raumolin-Brunberg (2003) の Corpus of Early English Correspondence に基づく社会階層別の三人称単数 *s* の使用頻度とその変遷を紹介している。

第 10 章は *web* をコーパスに見立てて活用する ‘Web as Corpus’ と, 新たなコーパスを構築のために *Web* を活用する ‘Web for Corpus’ について論じられている。コーパスとしての *Web* の問題点については, 赤野 (2007) がすでに指摘しているが, 著者はサーチエンジンが効率的な情報検索のためのものであって言語研究の検索に適していないこと, タグ付けされていないこと, テキスト属性の情報が限られていることなどの不備をあげつつ, コーパスとして活用する方法を紹介する。その例として, Google の Advanced Search の言語学的に有効な使い方, KWICfinder や WebCorp のような *Web* ページのきめ細かな検索と後処理ができるサイトの使用方法, ドメインごとに検索することで英語の変種間の差を調査するケースが示され, 研究事例として, 単音節の形容詞の比較変化に *Web* を活用した Mondorf (2007), ブログやフォーラムの英語を対象に非

標準的な動詞の過去形 (e.g. *She sung her song.*) を調査した Anderwald (2007) が紹介されている。

Web for Corpus の試みとして、携帯電話のメール、ブログ、Web 上のチャット、フォーラムなど話し言葉と書き言葉が混合した “a hybrid mode” の言葉を収集しコーパス化する動きや、Mark Davies の Time Corpus と COCA, Hundt and Biewer (2007) の South and East Asia Corpus 構築が報告されている。

以上、本書の内容を簡単に見てきた。200 ページそこそこの分量で、コーパス言語学で扱われるほとんどの基本概念が解説され、Web 上で利用できるコーパスに基づき英語の記述研究がどこまで可能かということ具体的かつ詳細に提示している本書は、大学上級レベルおよび大学院用の教科書にうってつけである。冒頭で述べたように、Study questions と hands-on exercises が用意されているので自習教材としても利用できる。さらに英語学を研究しているが、コーパスには不案内な研究者にも強く推薦できる良書である。

参考文献 (紹介順)

- Lindquist, H. & C. Mair (eds.) (2004), *Corpus Approaches to Grammaticalization in English*, Benjamins.
- Svartvik, J. (1992), ‘Corpus linguistics comes of age’, in Jan Svartvik (ed.), *Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82*, Stockholm, 408 August 1991, de Gruyter, 7–13.
- Procter, P. (ed.) (1978), *Longman Dictionary of Contemporary English*, Harlow and London: Longman. [LDOCE1]
- Summers, D. (ed.) (1995), *Longman Dictionary of Contemporary English*, Harlow: Longman. [LDOCE3]
- Firth, J. R. (1957), ‘A synopsis of linguistic theory, 1930–1955’, in *Studies in Linguistic Analysis*, special volume, Philological Society, Oxford: Blackwell, 1–32. Reprinted in F. R. Palmer (ed.), *Selected Papers of F. R. Firth 1952–59*, London: Longman, 168–205.
- Palmer, H. E. [1933] (1966), *Second Interim Report on English Collocations*, Tokyo: Kaitakusha.
- Altenberg, B. (1998), ‘On the phraseology of spoken English: The evidence of recurrent word combinations’, in A.P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications*, Oxford University Press, 101–122.
- Mahlberg, M. (2007), ‘Corpus stylistics: Bridging the gap between linguistics and literary studies’, in M. Hoey et al. (eds.), *Text, Discourse and Corpora*, Continuum, 219–246.
- Charteris-Black, J. (2004), *Corpus Approaches to Critical Metaphor Analysis*, Palgrave.
- Rudanko, J. (2006), ‘Watching English change: A case study on complement selection in British and American English’, *English Language and Linguistics* 10 (1), 31–48.
- Rissanen, M. (2004), ‘Grammaticalisation from side to side: On the development of beside(s)’, in Lindquist & Mair (eds.), 151–170.
- Mair, C. (2004), ‘Corpus linguistics and grammaticalization theory: Statistics, frequencies, and beyond’, in Lindquist & Mair (eds.), 121–150.
- Nevalainen, T. & H. Raumolin-Brunberg (2003), *Historical Sociolinguistics*, London: Longman.
- 赤野一郎 (2007), 「英語研究とネット活用」『月刊言語』36 (7), 50–57.
- Mondorf, B. (2007), ‘Reclacitrant problems of comparative alternation in newspaper CD as corpora’, in Kirk (ed.), *Corpora Galore: Analyses and Techniques in Describing English*, Rodopi, 57–71.
- Anderwald, L. (2007), ‘“He rung the bell” and “she drunk ale”: Non-standard past tense forms in traditional British dialects on the Internet’, in Hundt, et al. (eds.) *Corpus Linguistics and the Web*, Rodopi, 271–285.
- Hundt, M. & C. Biewer (2007), ‘The dynamics of inner and outer circle varieties in the South Pacific and East Asia’, in Hundt, et al. (eds.), 249–269.

本書で使用されているコーパス

- British National Corpus (BNC)
[BYU-BNC <http://corpus.byu.edu/bnc/>]
- Corpus of Contemporary American English (COCA)
[<http://www.americancorpus.org/>]
- Michigan Corpus of Academic Spoken English (MICASE)
[<http://quod.lib.umich.edu/m/micase/>]
- Phrases in English (PIE) [<http://pie.usna.edu/>]
- Time Corpus [<http://corpus.byu.edu/time/>]
- WebCorp Live [<http://www.webcorp.org.uk/>]

英語コーパス学会 Newsletter No. 70

Sept. 1, 2010

■会長: 赤野 一郎
■事務局: 〒739-8521 東広島市鏡山 1-7-1 広島大学大学院総合科学研究科 井上永幸研究室内
■TEL: 082-424-6431 ■振替口座: 00940-5-250586 (英語コーパス学会)
■URL: <http://muse.doshisha.ac.jp/JAECS/index.html> ■email: inoue@v.email.ne.jp

JAECS
Japan Association for English Corpus Studies

第 34 回大会のご案内

英語コーパス学会第 36 回大会は、10 月 9 日 (土)、東京大学 (駒場キャンパス; 〒153-8902 東京都目黒区駒場 3-8-1; http://www.u-tokyo.ac.jp/index_j.html) で開催されます。会場となる東大駒場キャンパスは、京王井の頭線 [JR 山の手線渋谷駅で乗り換え] で、吉祥寺方面行に乗って駒場東大前駅下車すぐのところにあります。事前に経路と時刻をご確認ください。宿泊を必要とされる方は早めの予約が肝要かと思われま。なお会場校の責任者である大堀壽夫先生には準備段階からご尽力いただき深く感謝致します。

詳細は、同封の「大会資料」をご覧くださいのですが、恒例になりました午前中のワークショップのほか、午後には研究発表 4 件とシンポジウムがあります。

午前中のワークショップでは、林良彦 (大阪大学) 氏に、「NLTK を用いた自然言語処理入門」と題して、自然言語処理 (NLP: Natural Language Processing) の教育用のプラットフォームとして注目されている NLTK (Natural Language Tool Kit; <http://www.nltk.org/>) を用いて、自然言語処理の入門的な紹介を行っていただきます。

研究発表については、運営委員の査読を経て準備委員会最終的な審査を行った結果、次の 4 件の発表が決まりました。今野昌俊 (東北大学大学院生) 氏の「英語の直接話法伝達部における句動詞の倒置」、仁科恭徳 (大阪市立大学非常勤講師)・平田恵理 (福岡女学院)・鈴木大介 (京都大学大学院生) 氏らの「Fair enough の辞書記述の改善案 - インフォーマント調査とコーパスの分析を通して -」、Warren Tang (広島大学大学院生) 氏の「Linearity and the Pedagogic Corpus: A Theoretical Study」、阿部真理子 (高崎経済大学) 氏の「コーパスを用いた日本人英語学習者の習熟度の分析」で、2 室に分かれての研究発表となります。

シンポジウムは、林良彦 (大阪大学) 氏の司会の下、

「コーパスにおける言語的注釈と自然言語処理技術」と題して、主として工学的な自然言語処理・言語資源管理の立場から、コーパスにおける言語的注釈に関するいくつかの話題を提供していただきます。まず、司会の林良彦氏に「言語的注釈データの生成と利用」と題して発表いただき、次に、永田昌明 (NTT コミュニケーション科学基礎研究所) 氏が「コーパスに基づく言語処理技術」、狩野芳伸 (東京大学) 氏が「自然言語処理プラットフォームの現状と利用」、乾健太郎 (東北大学) 氏が「コーパスへの意味的注釈の重層的付与」のタイトルでそれぞれ発表されます。

以上のように、今回は、NLTK を用いた自然言語処理に関するワークショップ、それに語法研究、英語教育を扱った 4 件の研究発表、さらにコーパスへのより良い注釈付けをテーマとしたシンポジウムと盛りだくさんの内容となっておりますので、多数のご参加を期待しております。

午前中のワークショップに参加ご希望の方は郵便・電子メール (件名「ワークショップ申込」、宛先は inoue@v.email.ne.jp) で、所属と会員・非会員の別を明記の上、事務局までお申し込みください。英語コーパス学会の会員であれば参加費は無料です (非会員の場合は当日会費 1,000 円)。

会誌『英語コーパス研究』第 18 号について

次回刊行の『英語コーパス研究』第 18 号への投稿を 9 月 30 日締め切りで受け付けております。たくさんのご応募をお待ちしております。現時点で、論文 4 点の投稿申し込みを受けております。

【原稿の種類】

1. 英語コーパス利用・コンピュータ利用を中心に据えた「研究論文」、「研究ノート」、「実践報告」。
2. 「書評」、「コーパス紹介」、「ソフト紹介」、「海外レポート」、「論文紹介」などの各種情報ある

いは紹介原稿。

【原稿提出締め切り】2010年9月30日(木)

※ハードコピー4部を郵送。およびWORDファイルを電子メール添付にて提出(電子メール件名には「英語コーパス研究18号投稿論文」とし、メール本体には、氏名、所属、原稿の種類とタイトルを書いて下さい)。論文、研究ノートの冒頭には題名のみ記し、氏名(ふりがな)、所属・職名、住所、電話番号、電子メールのアドレスを明記した別紙(1部)を添付のこと(WORDファイルもこの形式に沿ってください)。

【問い合わせ先・原稿提出先】

〒980-8576 仙台市青葉区川内41
東北大学大学院国際文化研究科 岡田 毅
TEL: 022-795-7632 FAX: 022-795-7632
Email: t-okada@intcul.tohoku.ac.jp

【原稿の長さ】

1. 研究論文
和文 A4サイズ1ページあたり35字×30行、17枚以内。
英文 A4サイズ1ページあたり70ストローク×35行、17枚以内(10.5ポイント使用)。
※いずれもAbstract(英文)、図表、注、書誌、付録を含む。
2. 研究ノート
1ページあたりは上記の書式と同様で、12枚以内。
※いずれもAbstract(英文)、図表、注、書誌、付録を含む。
3. その他
研究論文の半分以下。

【書式】

第17号所収の論文を参考にしてください。詳細は学会のWebページ(http://muse.doshisha.ac.jp/JAECs/Guidelines/ECS_SGuide-j.html)でご確認ください。

【採用通知】2010年11月頃

【刊行予定】2011年5月

『英語コーパス研究』編集委員会委員長
岡田 毅(東北大学)

新入会員紹介(8月25日現在、Sは学生)

北村 まゆみ	名古屋大学大学院 S
神戸 崇寛	成城大学大学院 S
今野 昌俊	東北大学大学院 S
西尾 美由紀	近畿大学
橋本 直樹	東京家政大学
Tang Warren	広島大学 S

事務局から

◇会費納入のお願い

2010年度会費(一般5,000円、学生3,000円)を、日本郵便にある払込取扱票を使ってお納めいただきますよう、ご協力をお願いいたします[振替口座:00940-5-250586]。日本郵便発行の受領証をもって領収書に代えさせていただきますので、ご了承ください。別途領収書が必要な方は、80円切手を同封の上、石川保茂(〒615-8558 京都市右京区西院笠目町6 京都外国語大学)までお申し出ください。払込取扱票の通信欄によるお申し出はご遠慮ください。

過年度会費未納の方は、2010年度分と併せてお納めください。会誌『英語コーパス研究』第17号は2009年度の会費を納入していただいた方にのみ、送付しております。また、2年続けて会費未納の場合、Newsletterなどの送付を中止させていただきます。

住所、所属などに変更や異動のある方は、必ず払込取扱票の通信欄にお書き添えください。

※会員の皆様には、日頃より会費の当該年度内納入のご協力をいただきまして、お礼申し上げます。会費を滞納されますと、退会時に滞納分をまとめてお支払いいただくといった事態にもなりかねません。会員の皆様におかれましては、円滑な学会運営のためにご協力いただけましたら幸いです。

◇メーリングリストについて

英語コーパス学会ではメーリングリストを使って会員の皆様の様々な情報交換に役立てていただいているところですが、最近、宛先不明でエラーが返ってくる例も増えています。会員の皆様方には、メールアドレスに変更が生じた場合、速やかに事務局宛ご連絡いただけますようお願い申し上げます。

FORUM

■ TaLC9 に参加して

中條清美 (日本大学)

chuujuu.kiyomi@nihon-u.ac.jp

2010年6月30日から7月4日にかけてチェコ第2の街、ブルノ (Brno) にある Masaryk University で開催された TaLC9 (9th Teaching and Language Corpora Conference) に参加してきました。TaLC は2年に1回開催されるコーパスの専門学会で、コーパスを言語教育に応用することに関心のある人々が集います。友好的で温かな雰囲気の中、コーパスの分野で著名な研究者の発表や意見を直接聞けるところが魅力です。英語コーパス学会からは鈴木千鶴子先生、横山彰三先生、安浪誠祐先生、投野由紀夫先生、中條が参加しました。キーノートが6講演、研究発表が50件、ワークショップが4件行われました。

以前に、TaLC で John Osborne (サヴォイ大学) が「コーパスの教育への応用は容易でない」と言っていました。実際に過去数回の TaLC ではコーパスの教育への利用研究の発表はあまり多くありませんでした。ところが、今回の TaLC9 では、DDL (data-driven learning) などコーパスの教育への直接利用の発表が15/50件と目立って多くなった印象がありました。私は自分も関心の高い DDL 関連の研究を中心に発表を聞きました。時間の関係で全てを聞くことはできなかったのですが、いくつか印象に残った研究を紹介します。

キーノート講演者の Ana Frankenberg (リスボン大学) は “Using corpora routinely” というタイトルで BNC, COCA, COMPARA (英語・スペイン語パラレルコーパス) を用いた翻訳指導の実例を紹介しました。その中で「コーパス vs. 辞書」ではなく、辞書の方が便利な時には辞書を使いましょうという姿勢が新しい視点でした。コーパスと辞書に関連する発表では、Claire Kennedy & Tiziana Miceli (グリフィス大学) は、3名のイタリア語学習者がコーパスと辞書を選択使用する時の目的と好みの傾向を詳細に分析して報告しました。今後、多様な言語の学習者を対象としてこのような研究が増えると面白いと思いました。

DDL についてはコーパスの「直接利用 vs. 間接利用」

という議論がありますが、この点に焦点をあてた実践報告もありました。Alex Boulton は “Hands-on, hands-off” という発表の中で、また、Chujo & Oghigian は “DDL for EFL beginners: recent gains and student views on the role of L1 and paper-based concordancing” という発表において、両研究とも同様の結果を報告しました。すなわち、computer-based DDL も paper-based DDL もそれぞれ長所・短所があるので、教育環境や学習者のタイプによって、どちらかを選択、あるいは両者を適切に組み合わせるのがよいであろうという妥当な結論に収束しました。

私がこれまでに参加した TaLC の中で印象に残っていることは、TaLC8 (2008 リスボン) でキーノート講演者のお一人の投野由紀夫先生が “TaLC in action: recent innovations in corpus-based English language teaching in Japan” という講演で NHK 『100語でスタート! 英会話』を紹介され、日本では「コーパス君」のおかげで「コーパス」が一般の人々にも広く知られているということが TaLC 参加者に驚きを与えたことです。以後、Guy Aston 氏 (ボロニア大学) は「Yukio は TV スターだ」と言い続けておられます。

今回と過去の大会の詳細は <http://tal2010.muni.cz/> より見ることができます。次回の TaLC10 は2012年で、開催地はまだ決定していないようですが、ぜひ参加したいと思っています。

■ PALA (Poetics and Linguistics Association) 年次大会参加報告

新井恭子 (東洋大学)

arai-k@toyo.jp

PALA は文体論と詩論、それに関連する学問領域、言語学、哲学、心理学、認知科学など、様々な分野の研究者から組織される国際的学術協会である。会員の居住地は50か国におよび、毎年、欧米のみならず、世界中の大学で年次大会が開催される。最近の発表の研究テーマは、narratology, literariness, literary linguistics, stylistics and pedagogy, critical discourse analysis, gender and writing, literary translation studies linguistics and philosophy, metaphor, cognition, pragmatics など多岐に渡る。

研究発表は和やかな雰囲気の中で行われ、同じ専門の研究者だけではなく異なった分野の研究者からの質

問やアドバイスもあり、学問領域を超えて新しい研究の見地が見いだされることも多い。また、各分野の著名な研究者の講演や発表を聞くことができるだけでなく、会期中に直接質問や相談もすることができ、若手研究者にとっても、非常に貴重な学会体験となっている。また、大会では、コーパスに関する最新技術の紹介やハンズオンのワークショップも開催され、会員には日本英語コーパス学会会員も多い。

集中した研究発表会の後に、主催者が趣向を凝らし実施する懇親会や観光ツアーは、開催国・地域の歴史や文化を楽しみながら学ぶことができ、この学会の魅力ともなっている。

今年の年次大会「PALA2010」は7月20日～25日、イタリアのジェノバ大学主催で開催され、統一テーマは、“The Language of Landscapes”であった。参加者は169人、内日本人が21人であった。この日本人の参加者数はイギリス人に継ぐ第2位であった。以下は今回の日本人・英語コーパス学会会員の発表である。(アルファベット順)：

Fuami, Shigenobu (Osaka Ohtani Univ.) “Wider Possibilities of Meaning in Early Modern English: Text, Collection and Interpretation,” **Hori, Masahiro** (Kumamoto Gakuen Univ.) “Development of Idioms in Dickens: With Special Referent to the Idiom ‘Line of Business,’” **Imahayashi, Osamu** (Hiroshima Univ.) “Description on Grammar and Usage in the Dickens Lexicon,” **Nishio, Miyuki** (Kinki Univ.) “The Development of Idiomatic Expressions in Dickens,” **Sera, Haruko** (Univ. of Hyogo) “Is business war in Japan? War metaphors in Michael Crichton’s *Rising Sun*,” **Tabata, Tomoji** (Univ. of Osaka) “A Multivariate Approach to Linguistic Variations in the Century



学会最終日、Conference Dinnerでの日本人参加者のテーブルにて。英語コーパス学会員は、向かってテーブル手前空席部分左より時計回りに、堀正広先生、お一人おいて、西尾美由紀先生、お一人おいて、今林修先生、お一人おいて、田畑智司先生が写っている。空席は筆者。[ジェノバ市内のレストラン“LE TERRAZZE DEL DUCAL”]

of Prose Corpus Part B: An Experiment in Corpus Stylistics” **Tabata, Tomoji** (Univ. of Osaka) and **Miyake, Maki** (Univ. of Osaka) “Mining the Landscape of Stylistics: A Multivariate Analysis of PALA Papers,” そして本報告筆者 **Arai, Kyoko** (Toyo Univ.) “‘Persuade to See the Same Landscape’—Revision of Cognitive Environment.”

中でも、The Dickens Lexiconの開発についての発表(広島大学の今林先生と近畿大学の西尾先生)また、その一部を利用した研究発表(熊本学園大学の堀先生)はたいへん興味深かった。入力されている情報量が多く、検索方法が簡単で、研究に有用な様々なリンクが張られており、完成すれば、ディケンズの研究のみならず、英文学教育にも非常に役立つ画期的なコーパスになるであろう。

次年度は、the Polytechnic of Namibia, Windhoek /vɪnt-hɔk/, Namibia で7月5～9日開催予定である。

※ PALA と PALA2010 については、URL: <http://www.pala.ac.uk/> を参照。

[この報告については、大阪大谷大学浮網茂信先生にご協力をいただいた]

■ Mac OS X 用コンコーダンサー CasualConc

田畑 智司 (大阪大学)

tabata@lang.osaka-u.ac.jp

1 CasualConc とは？

CasualConc は、今尾康裕氏 (カリフォルニア大学ロサンゼルス校博士後期課程修了) 作成の Mac OS X 専用多機能コンコーダンサーである。英語だけでなく日本語も扱うことが可能で、KWIC コンコーダンスの作成 (Concord) はもとより、Word Cluster 検索、コロケーションの検索や視覚化、語彙頻度・n-gram 計算、さらに、詳細なコーパスファイル情報を提示する機能などを備えている。執筆時現在、CasualConc はバージョン 1.0.3 である。専用のウェブサイト (<http://sites.google.com/site/casualconcj/>) からダウンロード可能なフリーウェアとして公開されている。さらに、このバージョンに加え、機能を強化した開発中の次期バージョン (CasualConc beta 1.7) もテスト使用可能である。CasualConc の動作環境は Mac OS X 10.5 (Leopard) 以降のバージョンであるが、作者によれば、Mac OS X 10.6.4 以降では OS の仕様変更のため、CasualConc beta の方が安定して動作するとのことなので、本稿は最新ベータ版 1.7 についての紹介とする。紙幅の都合上、ここでは詳細な使用方法や Tips について解説することはしないが、上記ウェブサイトには、スクリーンショットを数多く交えてこのプログラムの使い方が詳述されているほか、ディスカッションボード、同じ作者による他の処理ツールへのリンクなどが用意されているので、ぜひウェブサイトにアクセスしていただきたい。

2 分析対象のテキスト・フォーマット等

読者になじみが深いであろう AntConc、WordSmith Tools、TXTANA、KWIC Concordance for Windows 等、よく知られたコンコーダンサー同様、CasualConc はテキスト主体の文書を分析するためのツールである。処理に最適なのはプレーンテキスト (.txt) だが、XML、HTML、MS Word (.doc/.docx)、リッチテキスト (.rtf/.rtfd)、PDF などのファイル形式にも対応している。

CasualConc は、デフォルトでは段落 (改行文字で区切られた文字列) を文脈の範囲 (Scope of context) として分析を行う仕様になっている。検索対象の文脈語 (context words) や並べ替え語 (sort words)、クラスタ

(clusters) などの指定を行う際、段落を文脈の範囲とするため、分析対象のテキストは段落ごとに改行を施した形式で保存しておくのが最適である。そうすることで、コロケーションやクラスタの検索漏れを防ぐこともできる。(第 35 回大会で、文中の改行文字の存在などが検索結果に影響することを警告された大名力氏の講演はまだ記憶に新しい。) もちろん、細かなオプション設定を施すことにより、センテンスあるいはファイル全体を文脈の範囲として分析することもできる。ただし、そのトレードオフとして処理速度の低下を伴うかもしれない。センテンス単位の処理を最適化するには、一行一文のフォーマットでテキストを作成しておくとういだろう。

3 コーパスファイルの扱い

CasualConc のコーパスファイルの扱いには、File モードと、読み込んだファイルを SQL データベースとして管理する Database モードの二種類がある。それぞれに、基本となる Simple モードに加え、複数のファイルやデータベースをグループ化して扱う Advanced モードが用意されているので、計 4 つの様式から自分の用途に合ったファイル管理の方法が選択できるのは便利である。

はじめて CasualConc を起動する場合、File モード (Simple) の Concord ツールでテキストファイルを読み込むのがもっとも簡単な使い方であろう。ファイルの文字エンコーディングが UTF-8 もしくは ASCII の場合は、ファイルを読み込んですぐに分析が開始できる。その他のエンコーディングで保存されたテキストを分析する場合は、Default encoding を変更する必要がある。その他、マークアップ、アノテーションの施されたテキストやある程度のカスタマイゼーションを必要とする場合は、環境設定 (Preferences) の然るべき変更が必要になる。いずれにせよ、コーパスから正しい情報を引き出すには、分析ツールの環境設定を適切に行うことが不可欠であるので、利用者にはウェブサイト記載の「使い方」の各項目を熟読することをお奨めする。

4 XML 対応

CasualConc の強力な特徴の一つは、コーパスファイルに埋め込まれた XML タグの扱いである。最近では、XML によるコーパスファイルのアノテーションが積極的に進められている一方で、XML に本格的に対応し、

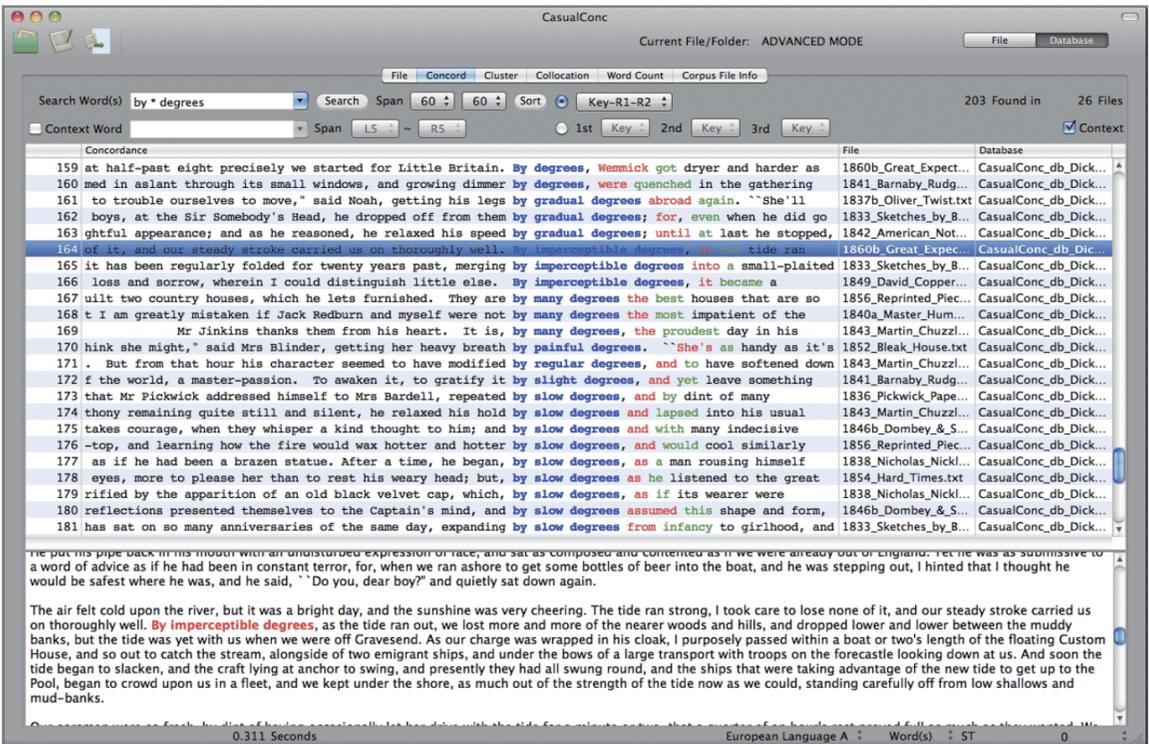


図 1 CasualConc: Concord ツール

なおかつ user-friendly なコンコーダンスツールはまだ少ないというのが現状である。Windows 用には、塚本聡氏による KWIC Concordance for Windows があるものの、Mac OS 用となると、コンコーダンスそのものが少ないこともあり、XML 情報タグ処理機能を備えたこのプログラムの存在は貴重である。

XML 情報タグを扱うには、環境設定 (Preferences) を変更する。まずメニューから、Preferences → Files → File Types の欄へ移動し、 XML (.xml) にチェックマークを入れた後、さらに、Preferences → XML へ移動する。ここで、 XML File Information にもチェックマークを入れた後、XML ヘッダーの形式を指定する。XML 情報タグのうち特定の要素をフィルターとして検索対象を限定することにより、コーパスファイル中の特定の構成要素、テキストタイプ、著者・話者などに対象を絞った検索を行うことが可能になる。例えば、小説の地の文 (narrative) と会話部を区別するマークアップを施していれば、narrative のみ、会話部のみに絞った語彙・n-gram 検索やコンコーダンス作成、コロケーションの検索が行える。

5 ツール

CasualConc は、大きく分けて Concord, Cluster, Collocation, Word Count, Corpus File Info の 5 つのツールからなる。以下、順を追って簡単な説明を行う。

CasualConc のメイン機能である Concord [図 1] は、英語、仏、独、西語などの西欧諸語に加え、ギリシャ語のように特殊なアルファベットを用いる言語も扱えるほか、日本語のテキスト (分かち書き有・無) にも対応している。

KWIC コンコーダンス検索では、全用例のうちから標本数を指定し、無作為抽出することができる。他方、検索語 (keyword) に加えて、文脈語を指定することにより、特定の文脈語を (含む | 含まない) 用例に絞り込む機能もサポートしている。その他、面白い仕様としては、検索語を括弧と空白に変えて表示させるオプションも備えている。この機能は、類似の文脈から類推して空所補充をさせるような問題を作成する時など、教材開発の場面でも活用できるのではないだろうか。

単語クラスターを検索する Cluster ツールは、検索語を含め 2 ~ 5 語の単語クラスターを検索できる。検索語は 1 語とカウントされるので、もし検索語入力欄に

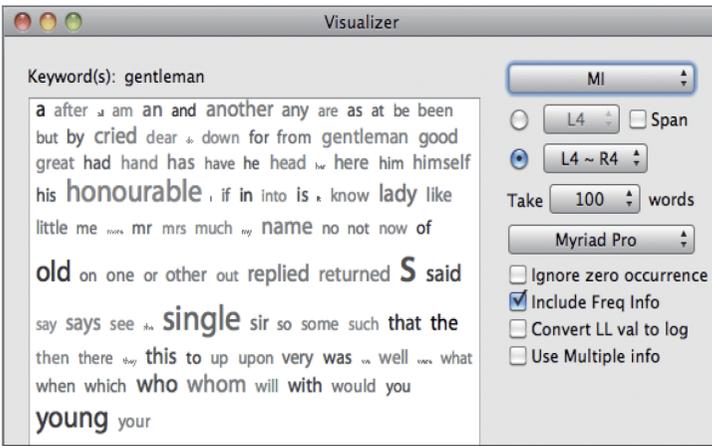


図 2 Collocation Visualizer

“as if she”のように3語を記入すれば4～7語の単語クラスターを検索することも可能である。単語クラスターの検索では、検索語の生起位置を特定しない設定 (Span = All) の他、他の語が検索語の (右側 | 左側) にのみ生起する用例に限定することもできる。単語クラスターの検索結果を表示するテーブルは2つ用意されているので、左右で異なる語数のクラスターを表示させて検索結果を見比べることができるようになっている。

共起語検索や共起語テーブルの作成を行う

Collocation ツールも機能

が充実している。検索語を

Search word(s) 欄に入力し、

Search ボタンをクリックすると、

検索語を中心に左右5語の

スパンに生起する語彙の共起頻度

テーブルが表示される。テーブルは

共起の総頻度のほか、検索語の

(左側 | 右側) のみの頻度、生起位置ごとの

頻度などの基準で並び替えをすることが

できる。

また、Cooccurrence テー

ブルの表示機能も実装

されており、Collocation

ツールのウィンドウの下側に

位置している Cooccurrence

タブをク

リックすると、左右5語および検索語がそれぞれの位置で頻度の多い順に表示される。

コロケーションの分析においては、検索語と共起語との結びつきの強さを表す統計量を適切に解釈することが重要である。CasualConcでは実頻度 (Raw frequency) の他、MI Score, MI3 Score, Log-Likelihood ratio, z-Score, t-Score, LogLog などの統計指標を用いてコロケーションの強度を示すことができる。さらに、応用的な使用方法として、Stats メニュー

の下側に配置されている Collocation Stats Calculator や

Contingency Table Calculator などの計算ツールを使うこと

で、共起統計量の計算や、(一般のコンコordanサーの用途を超えるが) 分割表作成と尤度比検定、フィッ

シャーの正確確率検定等を行うことも可能である。

Collocation ツールの紹介の最後に、ユニークな特徴【コロケーションを視覚化する機能 (Collocation Visualizer)】

[図 2] について触れておきたい。この機能はまだベータ版での実装試験中であるため、日本語版のウェブサ

イトにはまだ記載がない。現時点では、英語版のウエ

Words	Frequency	In Files	Proportion	Log-Likelihood	Words	Frequency	In Files	Proportion
1 the	247344	28	5.11%	2019.344	1 the	237550	31	4.49%
2 and	176689	28	3.65%	2409.402	2 to	164845	31	3.12%
3 of	131644	28	2.72%	90.737	3 and	163142	31	3.08%
4 to	127324	28	2.63%	-2078.752	4 of	138673	31	2.62%
5 a	109111	28	2.26%	715.354	5 a	106453	31	2.01%
6 in	85455	28	1.76%	406.727	6 i	97207	31	1.84%
7 i	81727	28	1.69%	-319.803	7 in	84679	31	1.60%
8 that	63581	28	1.31%	-198.509	8 was	77734	31	1.47%
9 his	61408	28	1.27%	1301.717	9 that	74955	31	1.42%
10 he	61006	28	1.26%	-230.532	10 her	72566	31	1.37%
11 it	59861	28	1.24%	94.353	11 he	72457	31	1.37%
12 was	56074	28	1.16%	-1863.851	12 it	61868	31	1.17%
13 you	49428	28	1.02%	-0.214	13 she	56723	31	1.07%
14 with	47414	28	0.98%	612.060	14 had	56635	31	1.07%
15 as	41502	28	0.86%	-371.772	15 his	54267	31	1.03%
16 had	35944	28	0.74%	-3010.728	16 you	54167	31	1.02%
17 for	34865	28	0.72%	-347.219	17 as	51489	31	0.97%
18 her	34406	28	0.71%	-10743.52	18 not	45942	31	0.87%
19 at	34082	28	0.70%	12.733	19 with	43987	31	0.83%
20 mr	31312	28	0.65%	3436.603	20 for	43545	31	0.82%
21 said	30701	28	0.63%	725.849	21 be	41134	31	0.78%
22 on	30162	28	0.62%	4.668	22 but	38452	31	0.73%
23 is	28395	28	0.59%	51.887	23 at	36253	31	0.69%
24 him	27828	28	0.57%	-0.035	24 on	32394	31	0.61%
25 my	27786	28	0.57%	-0.284	25 have	31396	31	0.59%
26 have	27318	28	0.56%	-37.227	26 my	30497	31	0.58%
27 be	26788	28	0.55%	-1914.301	27 him	30456	31	0.58%
28 but	26202	28	0.54%	-1375.848	28 is	29220	31	0.55%
29 not	26200	28	0.54%	-3863.108	29 said	26791	31	0.51%
30 me	22567	28	0.47%	-57.832	30 me	26418	31	0.50%
31 this	22338	28	0.46%	467.832	31 which	24260	31	0.46%
32 by	22233	28	0.46%	128.664	32 would	22961	31	0.43%
33 she	22142	28	0.46%	-12779.23	33 so	22357	31	0.42%
34 so	20986	28	0.43%	6.979	34 all	22086	31	0.42%

図 3 Word Count

ブサイトにのみ使用法が解説されている。視覚化の手順はいたって簡単で、共起語テーブルを作成した後、Visual ボタンをクリックするだけである。これにより、各統計指標に基づく共起の強さを図示することができる。例えば、ディケンズの全文コーパスにおける gentleman の共起語データは図 2 のようなダイアグラムで視覚化される。この例では文字の濃淡が gentleman の共起語の実頻度を示す一方、文字サイズの大小に MI Score を反映させている。このように共起の特徴を視覚的に表示する機能は有益であろう。

Word Count ツール [図 3] では語彙や n-gram のリスト作成が行える。検索結果は単語クラスターの場合と同じように二つのウィンドウに表示可能なので、左右で異なるコーパス (ファイル) の語彙頻度リストを表示して比較分析ができる。また、二つのコーパス (ファイル) から得られた頻度表をもとに、語の keyness を計算することも可能である。図 3 は左半分にディケンズの全文コーパスの語彙頻度リストを、右半分に他の 19 世紀英国小説コーパスの語彙頻度リストを表示している。左側の Stats のコラムには (ディケンズを基準にした) 語彙の keyness を示す統計量 (ここでは Log-likelihood ratio) が表示され、そのうち赤字 (モノクロ紙面では薄いグレー表示) のものは 'negative' key-words である。語彙・n-gram のリストは、頻度、アルファベット順、単語末尾からのアルファベット順、語長 (文字数)、統計値 (keyness) を基準にした並び替え表示もサポートしている。

Corpus File Info ツールの基本的機能は、コーパスファイル毎の type 数、token 数、type-token ratio (TTR)、単語の平均語長 (文字数)、1 ~ 15 文字の語および 16 文字以上の単語の数を表示することである。他方、応用的な使用方法として、各コーパスファイルごとの語彙頻度プロファイルや、TF-IDF 方式による語彙重み付けテーブルを作成するユニークな機能も実装されているので、テキストマイニングを行う上で大変重宝するツールである。

6 ワイルドカード検索

Concord, Cluster, Collocation ツールでの検索時には、正規表現を使った複雑な検索も行える。一方で、* (0 または 1 文字以上の文字列に適合)、! (1 文字に適合)、? (1 文字以上の文字 (列) に適合)、/ (or)、(A|B) (A or

B) などの記号を用いたワイルドカード検索もサポートしている。たとえば、by *degrees で検索を行えば、by degrees に加え、by (gradual | imperceptible | painful | regular | slight | slow) degrees などの用例を見つけることが可能である。

7 検索結果の保存、書き出し

KWIC コンコーダンスや共起語テーブル、語彙・n-gram リストなど、各ツールのアウトプットは、CasualConc フォーマットで保存できるほか、タブ区切りテキスト (.txt)、または、コンマ区切りデータ (.csv) の形式でファイルに書き出せるようになっているので、ワードプロセッサや表計算のプログラムに読み込むことも可能である。

8 最後に

これまで、CasualConc の基本機能と、可能な範囲でユニークな特徴を紹介してきたが、最後に、筆者からの期待と要望を挙げておきたい。まずは、CasualConc のさらなる操作性向上への期待である。Mac OS X 専用のツールとしては他に類を見ない高機能なプログラムであることを認めた上で、敢えて注文をつけるとしたら、それは各ツール間のシームレスな連携であろうか。たとえば、Cluster, Collocation, Word Count のウィンドウに表示されている語や句をクリックすることで、その語・句の KWIC コンコーダンスへ即座に移行できるようになれば、まさに鬼に金棒だと筆者は考える。可能ならば将来のバージョンでの実装に期待したい。次に、「要望」はユーザに対してである。現在の CasualConc の姿は、作者今尾氏によるさまざまな機能強化を目指した弛まぬ創意工夫、多大な時間と労力を尽くした成果である。ユーザからのバグ情報や要望などのフィードバックはプログラムの安定性・機能向上に不可欠なだけでなく、プログラミングを続ける大きな動機づけにもなるはずである。使用した感想や意見、要望などを、作者に積極的にフィードバックしていただければ幸いである。スパム防止の観点から作者の電子メールアドレスの記載はここでは差し控えるが、ウェブサイト記載のアドレスや Twitter のアカウント (@casualconc) 宛てにコンタクトをとるのも一つの方法だろう。

CasualConc is a super concordancer for the rest of us!

英語コーパス学会 Newsletter No. 71

Dec. 1, 2010

■会長: 赤野 一郎
■事務局: 〒739-8521 東広島市鏡山 1-7-1 広島大学大学院総合科学研究科 井上永幸研究室内
■TEL: 082-424-6431 ■振替口座: 00940-5-250586 (英語コーパス学会)
■URL: <http://muse.doshisha.ac.jp/JAECS/index.html> ■email: inoue@v.email.ne.jp

JAECS
Japan Association for English Corpus Studies

第 36 回大会報告

■概要

英語コーパス学会第 36 回大会は、10 月 9 日(土)、東京大学(駒場キャンパス)で開催されました。あいにくの雨にもかかわらず、京王井の頭線駒場東大前駅下車すぐという立地条件の良さもあって、101 名の参加がありました〔会員 72 名+新入会員 4 名+当日会員 25 名〕。

午前中のワークショップは「NLTK を用いた自然言語処理入門」と題して林良彦先生(大阪大学)に講師を務めていただきました。自然言語処理(NLP: Natural Language Processing)の教育用のプラットフォームとして注目されている NLTK (Natural Language Tool Kit; <http://www.nltk.org/>) を用いて、自然言語処理の入門的な紹介を行っていただきました。特に、英語コーパスの処理に関心がある方が主な聴衆であることを想定し、英語コーパスの基本的な処理、コーパス処理に有用なデータ型、単語の正規化処理、基本的な言語解析、品詞タグ付け、構文解析といった内容を含めて解説していただきました。当日は、可能な場合は各自ノート PC を持参するという設定でしたが、ノート PC を持参していない参加者も熱心にメモを取る姿が見られました。ワークショップの内容を反映し、39 名の参加者のうち、当日会員の参加者が多くを占めていたようです。参加者の皆様からも、非常に有意義なワークショップであったとの評価をいただいております。講師を務めていただきました林良彦先生には、この場をお借りしてお礼申し上げます。

午後の大会では、まず赤野一郎会長(京都外国語大学)の開会の挨拶があり、引き続き、学会賞選考委員長の投野由紀夫先生(東京外国語大学)から学会賞の発表と選考理由の説明の後、下記のような結果が発表されました。

第 10 回英語コーパス学会賞

学会賞

受賞者: 金子朝子氏(昭和女子大学)

受賞対象: 「ICLE/LINDSEI Japanese subcorpus の構築およびこれらを利用した学習者コーパスの一連の研究」

引き続き「言語研究」と「英語教育」のテーマで 2 室に分かれて、林裕先生(関東学院大学)と羽山恵先生(獨協大学)の司会のもと、それぞれ 2 件ずつの研究発表が行われました。

最後に、午前中のワークショップと連動して、林良彦先生の司会のもと、「コーパスにおける言語的注釈と自然言語処理技術」と題するテーマでシンポジウムが行われました。それぞれの司会の先生にご執筆願いました概要につきましては、「研究発表」及び「シンポジウム」のセクションをご覧ください。最後に、開催校の大堀壽夫先生(東京大学)の閉会の辞をもって、第 36 回大会の幕を閉じました。

大会終了後の懇親会は、45 名の出席がありました。堀田秀吾先生(明治大学)の軽妙な司会のもと、会長の赤野一郎先生の挨拶、そして会誌編集委員長である岡田毅先生(東北大学)の乾杯のご発声で始まりました。途中で、開催校を代表して嶋田正和先生(東京大学大学院総合文化研究科副研究科長)にご挨拶をいただき、懇親会の場がいつそう和んだものとなりました。会員同士の交流と情報交換で盛り上がり、午後 8 時にはすべての大会行事が終了いたしました。

最後に、本学会員で開催校責任者の大堀壽夫先生のご尽力と細部にまで配慮の行き届いたご協力で盛会に終わったことを喜び、心よりお礼申し上げます。加えて大会実施に協力くださった学生、院生の方々にもこの紙上を借りて厚くお礼申し上げます。

■研究発表

英語の直接話法伝達部における句動詞の倒置

今野昌俊（東北大学大学院生）

本発表では、英語の直接話法伝達部における句動詞の倒置に関して、特に文頭での倒置の生起に着目し、BYU-Time Magazine Corpus を中心に行なった調査と、インフォーマント調査の結果が述べられた。先行研究では伝達部の句動詞が go on である場合、倒置は容認されないとされてきた (Kahn 1985)。一方、それ以外の句動詞 (call up, shout out, chip in など) では容認可能な例が報告されている。

文頭での倒置が Time Magazine で特によく目につくという指摘があるので、BYU-Time Magazine Corpus による調査を、back, down, in, out, up, on を含む句動詞の倒置について行ない、文頭 176、文中 52、文末 84 という数値を得て、文頭での倒置という Time style の存在が確認された。なおこの中には許容されないとされる go on の倒置が 5 例含まれる。年代別での数値を見ると go on は 1920 年代と 1930 年代のみである。

イギリス英語、アメリカ英語の母語話者の各 10 人の調査が go on の倒置について行われ、容認される場合があることが確認された。先行文脈を考慮しない文レベルでの判断、先行文脈で話者交代がある場合、同一話者の発話が継続する場合、及び主語名詞句が重い場合の判断を比較すると、主語名詞句が重いと倒置が自然であると答える話者が多く、また文レベルと同一話者の発話の継続において、アメリカ英語よりもイギリス英語でのほうが若干容認度が高くなることが示された。

質疑応答ではコーパス調査とインフォーマント調査の結果の乖離がある場合について、年代別調査に関して最近公開された BYU-COHA の利用について、伝達部を手掛かりにした検索によるアプローチについて、自動詞・他動詞の区別の考慮についてのコメント・ディスカッションがあった。

林 裕（関東学院大学）

Fair enough の辞書記述の改善案：インフォーマント調査とコーパスの分析を通して

仁科恭徳（大阪市立大学非常勤）

平田恵理（福岡女学院大学）

鈴木大介（京都大学大学院生）

本発表では、英語の fair enough という表現につい

てコーパス分析とインフォーマント調査をおこない、そこで得た知見を基に辞書記述の改善案が提案された。

BNC と COCA で使用頻度を見ると、アメリカ英語よりもイギリス英語で約 4 倍程度多い。いずれにおいても七割台を話し言葉が占めることから、fair enough はダイアログ表現で使用されることが多い。また、1985 年から現在まで継続的に使われている現代英語表現であり、使用者年齢としては中高年層で好まれ、男女別の差はほとんど見られないと報告された。

13 名の英語母語話者をインフォーマントとして、fair enough 発話時の心的態度をアンケート調査した。これは fair enough と言う場合に、ダイアログの相手の発話内容に、「全く同意」、「部分的に反対」、「全く反対」、「その他」の 4 つのいずれ（複数回答可）の心的態度を持つのかを回答してもらい、選択した心的態度の具体的ダイアログの作成も要請するというものである。

その結果、「部分的に反対」を選んだ母語話者が 92% と圧倒的に多く、「全く反対」が 31% とそれ続き、「全く同意」は 23% に留まった。また「その他」には、相手の発話に触発されて「新発見」がある場合にも使われることが分かった。

以上を踏まえて、fair enough の辞書記述に、「部分否定」「新発見」を加え、多く見られる使い方の順に、「部分否定」「肯定」「否定」「新発見」の順で掲載することを提案した。

活発な質疑応答があったがその一つを挙げれば、本発表での手順をシステム化、簡略化しマニュアルにすれば実際の辞書編纂に有益であるという指摘があった。

林 裕（関東学院大学）

Linearity and the Pedagogic Corpus: A theoretical Study

Tang Warren（広島大学大学院生）

本発表は、言語教育に用いるコーパスに対する「直線性 (linearity)」という観点を検討し、その有効性を論じる理論研究である。言語教育用コーパス (pedagogic corpus) の一例として、リーディング学習に用いられる小説を挙げた上で、多くの事象 (小説の構成や読み進めること、教育の課程など) が「直線的に」進むことを踏まえ、新たなコーパス分

析方法論を提案している。

現在広く用いられているコーパスとその分析手法に対して発表者が抱いている問題意識は、テキスト間の関係性が欠けており、コーパス全体から結果を求める分析は分散的になりがちな点である。教材として用いられる小説などは、段落や章に順序性があり、またそれらは互に関連している。そのため直線性の概念が有効だと考え、その証明としてある小説に出現する単語の累積的カバレッジと、それにより期待される偶発的語彙学習を例に挙げている。つまり、ある単語がコーパス（小説）のどこで出現したかを特定しておけば、その後に出てくる同単語は既出（既習）であり、読み手にとっての未知語の割合を知ることが出来るということである。

質疑応答では、今回提案された「直線性」は従来のコーパス分析手法に取って代わるような paradigm shift ではなく、研究目的等に応じた一つの可能性ではないかとの指摘があった。また、今回一つの小説をコーパス例に挙げたことに対し、そもそもコーパスをどう定義するのかという問いかけもあり、興味深いやり取りがくり返された。

羽山 恵（獨協大学）

コーパスを用いた日本人学習者の習熟度の分析

阿部真理子（高崎経済大学）

本発表は、比較的大規模であり習熟度情報が付与されている日本人英語学習者の話し言葉コーパス、NICT-JLE コーパスを Biber (1988) で用いられている複数の言語項目によって分析した結果の報告である。学習者言語の分析は個別の言語項目に特化していることが多く、多面的な記述に至っていないというのが本研究の持つ問題意識である。

分析は、品詞タグの付与→言語項目の頻度算出→コレスポネンス分析の手順を踏んだ。得られた結果は、1) Mode, 2) Topic, 3) Reference, 4) Function の4つの観点から考察された。それによると、1) 初級者の発話は名詞中心で書き言葉的、上級者は動詞中心で話し言葉的な特徴を持つ。2) 初級者は人称代名詞と be-動詞、叙述形容詞を用いて個人的な状態について述べ、習熟度が上がるにつれて心情や考えを述べる場合に private verb を用いる可能性がある。3) 習熟度が上がるにつれ、繰り返しを避けたり前出事項を参照したりできるようになる。4) informational な働きをする言語項目と結びつきが強

いのは中級レベルである。さらに、学習者の習熟度を特徴づける項目は、Biber (1988) のうち negation, downtoners/hedges, coordination, personal/impersonal pronouns, pronouns/pro-verb *do* だと特定された。

質疑応答では、コレスポネンス分析を用いる妥当性、学習者言語の分析に Biber (1988) の項目を用いる妥当性が問われた。これに対して発表者からは、今後の研究ではクラスター分析なども合わせて採用する予定であること、同一データに対してさまざまな項目を用いて分析することによって多面的な記述が可能になるはずだとの考えが示された。

羽山 恵（獨協大学）

■シンポジウム

シンポジウム《コーパスにおける言語的注釈と自然言語処理技術》

Web コーパス概観

司会 林 良彦（大阪大学）

近年の自然言語処理技術では、大規模なコーパスを前提とするデータ駆動型の手法が主流となっている。機械学習の適用においては、適切に注釈付けがなされたコーパスが必要である。一方で、様々な言語現象に対して適切な注釈付けがなされたコーパスに基づくことにより、言語学的な知見が深まり、さらに良い注釈付けの枠組みが得られることが期待される。以上のように、自然言語処理技術と注釈付きコーパスは不可分の関係にあり、なかでもコーパスに付与される言語的注釈は、言語学の理論的研究と計算機による言語処理研究の接点となる。本シンポジウムでは主として工学的な立場からコーパスにおける言語的注釈に関する重要な話題を提供することにより、学際的な研究交流の一助となることを目的とした。

言語的注釈データの生成と利用

講師 林 良彦（大阪大学）

まず、データ駆動型の自然言語処理の大きな枠組が提示され、これを背景とした本シンポジウムの趣旨が紹介された。次に、質の良い言語的注釈を効率よく得るための“Annotation Science”の概念が示され、この文脈の中で、各講師のトピックの位置づけが提示された。さらに、コーパスなどの言語資源が幅広く利用されるためには、相互運用性

(interoperability) の確保が重要であることが述べられ、標準的なデータモデルである Linguistic Annotation Framework (LAF) と、これを具体的に注釈付きコーパスに適用した Open National American Corpus の概要が紹介された。

コーパスに基づく言語処理技術

講師 永田昌明 (NTT コミュニケーション
科学基礎研究所)

本発表ではまず、なぜデータ駆動型の自然言語処理技術が中心となってきたかという背景的な説明に続き、データ駆動型の処理の中心となる機械学習技術の流れが見通しよく示された。次にこの中から特に、隠れマルコフモデルによる品詞タグ付け、確率文脈自由文法による句構造解析、識別学習による依存構造解析技術に関して、その数理的な考え方を中心に解説が行われた。また、機械学習手法を大きく分類する「生成モデル」と「識別モデル」の違いについて、身近な例を用いた説明が行われるとともに、今後の機械学習適用の方向性が示された。

自然言語処理プラットフォームの現状と利用

講師 狩野芳伸 (東京大学)

自然言語処理技術の発展とともに、数多くの言語資源 (注釈付きコーパスや品詞解析器等の自然言語処理ツール) が一般に利用可能な形で公開されているが、これらの多くは相互運用性が考慮されておらず、言語資源を組み合わせるためには、データ形式の変換や実行設定記述などの作業が必要である。本発表では、このような状況を改善するための自然言語処理プラットフォームの意義が述べられ、さらに、狩野氏が中心となって開発を進めている、UIMA フレームワークによる U-Compare プラットフォームが実例やデモンストレーションを交えながら分かりやすく紹介された。

コーパスへの意味的注釈の重層的付与

講師 乾健太郎 (東北大学)

コーパスに基づく言語解析の研究の深化・進展に伴い、コーパスへの注釈付けも語義、固有表現、照応・共参照、述語項構造、モダリティ、時間情報、談話関係など、多様化が進んでいる。本発表では、こうした多様な意味的注釈付けに関する海外での研究動向が概観されるとともに、乾氏らのグループに

よる重層的な言語学的注釈付けの仕様の具体例が紹介された。注釈付けの仕様は、漠然とした言語解析という問題を具体的な課題に切り分けるという問題に密接に関連することから、言語学研究と自然言語処理研究のより密接な連携が呼びかけられた。

新入会員紹介 (11月20日現在、Sは学生)

秋葉 倫史	日本大学
今井 信義	福井県立羽水高等学校
奥 聡一郎	関東学院大学
田中 秀毅	広島女学院大学
仲本 康一郎	山梨大学
土方 裕子	東京理科大学
藤本 和子	創価大学
渡辺 浩平	筑波大学 S
Victoria Muchleisen	早稲田大学

第 37 回大会研究発表者募集

Newsletter No. 69 でもお知らせしましたように、2011 年度からは、これまで春と秋の年 2 回開催して参りました研究大会を、より充実したものにすべく、年 1 回 2 日間にわたる開催へと移行することになっております。記念すべき年 1 回開催の第 1 回目は、10 月 1 日 (土)、2 日 (日) に京都外国語大学で行われる運びとなりました。つきましては、発表を希望される方は、下記の要領に従って email で事務局宛にお申し込み下さい。

【分野】 本学会にふさわしい、コーパス利用・コンピュータ利用を中心に据えた英語研究。

【応募資格】 本学会員であること。

【発表方法】 発表 20 分、質疑 10 分。

【応募方法】 冒頭に題名のみを記し、800-1200 字 (参考文献は別) にまとめ、メール添付ファイルで送付。メール本文に氏名 (ふりがな)、所属・職名、住所、電話番号、メールアドレス明記。

※審査の際、応募者が特定されないよう、事務局が応募書類を加工させていただくことがございます。

【応募締め切り】 2011 年 6 月 15 日 (水) 必着

【採否決定】 2011 年 7 月下旬 (予定)

【問合せ】 〒739-8521 東広島市鏡山 1-7-1

広島大学大学院総合科学研究科

井上永幸研究室内

英語コーパス学会事務局

email: inoue@v.email.ne.jp

ハンドアウトのダウンロードサービス

第 35 回大会の研究発表とシンポジウムのハンドアウトを希望される会員に対して、ダウンロードのサービスを行います。期間は、このニューズレターお届けより 12 月 24 日までとします。ファイルは PDF となっております。ご希望の方は、石川保茂 (yasuishikawa@hotmail.com) まで下記のハンドアウトのうちご希望の番号をお知らせください。追って URL をお知らせいたします。

なお、発表者の著作権保護の立場から印刷は「許可しない」に設定してあります。
※以下、講師・発表者の敬称は略させていただきます。

- 1 林良彦「NLTK を用いた自然言語処理入門」
- 2 今野昌俊「英語の直接話法伝達部における句動詞の倒置」
- 3 仁科恭徳・平田恵理・鈴木大介「Fair enough の辞書記述の改善案 ―インフォーマント調査とコーパスの分析を通して―」
- 4 Tang Warren “Linearity and the Pedagogic Corpus: A Theoretical Study”
- 5 阿部真理子「コーパスを用いた日本人英語学習者の習熟度の分析」
- 6 林良彦「言語的注釈データの生成と利用」
- 7 永田昌明「コーパスに基づく言語処理技術」
- 8 狩野芳伸「自然言語処理プラットフォームの現状と利用」
- 9 乾健太郎「コーパスへの意味的注釈の重層的付与」

会誌『英語コーパス研究』第 18 号について

『英語コーパス研究』第 18 号 (2011 年刊行) について報告いたします。

2010 年 9 月投稿締め切りの本年の会誌への投稿数は、研究論文 9 点、研究ノート 2 点、シンポジウム論文 1 組でした。昨年の投稿数に近いほど多くの原稿が寄せられたことになり、編集委員会としましては嬉しい悲鳴をあげているところです。現在、論文査読委員により査読審査が行われております。

『英語コーパス研究』は例年 9 月末日が投稿締め切りです。本年度投稿に至らなかった論文等をお持

ちの会員の方々は、来年度の投稿へ向けて是非ともご検討をお願いいたします。研究論文、研究ノートのみならず、書評やコーパス紹介、ソフトウェアレビュー、実践報告なども受け付けております。学術的水準も高く、かつ多くの会員諸氏にとって有益な情報をもたらす会誌としたいと願っております。多くの投稿をお待ちしております。

『英語コーパス研究』編集委員会委員長
岡田 毅 (東北大学)

運営委員会の決定事項について

大会前日の 10 月 8 日 (土) 午後 5 時 30 分より開かれた運営委員会において以下の議案が審議されました。

■人事について

まず、次期副会長に堀正広先生 (熊本学園大学) が会長から指名され、承認されました [ただし、任期は、特例として 1 年とする]。これに伴い、山崎俊次先生 (大東文化大学) の退任が認められました。山崎先生、お疲れ様でした。事務局長の井上永幸 (広島大学)、会計の石川保茂先生 (京都外国語短期大学) については、特例として 2011 年度 1 年間に任期とすること、また、次期事務局長、次期会計を念頭においた事務局補佐、会計補佐を置くことが認められました。

次に運営委員ですが、発足以来長きにわたって当学会の発展に貢献してこられた中村純作先生の今年度限りでのご退任、その他の運営委員については、次期継続とすることが承認されました。中村先生のこれまでのご尽力に心よりお礼申し上げます。2 期にわたり監事を務められた梅咲敦子先生 (関西学院大学) は退任が認められ、後任を検討してゆきます。梅咲先生、お疲れ様でした。また、堀正広先生の次期副会長就任に伴い、編集委員会委員は退任され、後任推薦を依頼することとなります。

さらに、2010 年度任期満期の役員の後任についてです。学会賞選考委員会に関しては、2 期にわたって [うち、1 年は深谷輝彦先生が代行] 委員長を務められた投野由紀夫先生の退任が承認され、後任を今後検討してゆくこととなりました。また、今年度で任期が切れる選考委員の深谷輝彦先生 (椋山学園大学) については、継続とすることが認められました。投野先生には、心よりお礼申し上げます。

■「運営委員」の名称について

「運営委員」の名称を「理事」、「論文審査員」の名称を「論文審査委員」に変更することが承認されました。

■事務処理・手続きの変更について

(1) 年間スケジュールの見直しについて

研究大会の年1回化に伴い、春の運営委員会開催が認められました。ただし、運営委員の旅費は学会が支給し、5月の中旬あたりを目処に開催することが承認されました。

会誌の発送は6月1日発送とすることが承認されました。

Newsletterの発行は、6月1日と12月1日とすること、6月1日発送のNLには会誌を同封することが承認されました。

学会賞の応募締切は5月15日前後、採否決定は7月下旬とすることが承認されました。

研究発表に関しては、従来どおりとすること、2回のNLに研究発表募集を掲載することが承認されました。

(2) 今後の決算報告と予算案について

春に開催される運営委員会で審議、議決し、秋に開催される大会時の総会において、報告事項とすることが承認されました。

(3) 会則の改定について

以上のこれらの事務処理・手続きの変更に伴う会則の改定が承認されました。次期総会において、会員の皆様にお諮りする予定です。

英語コーパス学会賞募集

第10回英語コーパス学会賞を募集いたします。学会賞は英語コーパス学会の活性化のために設けられた賞です。今回から〆切が5月になりました。奮ってご応募ください。特に、これまでに学会誌『英語コーパス研究』に掲載経験のある若手研究者で以下の条件に合う会員の方は、奨励賞にぜひ挑戦してみてください。

【対象】英語コーパス学会の目的にてらし、英語のコーパス言語学に関する優れた研究業績をあげた学会員（個人またはグループ）とする。ただし、奨励賞は英語のコーパス言語学に関する優れた研究業績（論文の場合、学会誌『英語コーパス研究』に掲載されたものに限る）をあげた

35歳以下または大学院修了後の研究歴5年以下の学会員個人に限る。

【応募方法】自薦、他薦を問わない。

【提出書類】1) 同封の推薦理由書。2) 論文の場合は現物またはコピー。単行本の場合は事務局で用意するので送付は不要。

【提出先】事務局。

【応募期限】2011年5月15日（日）

【発表】2011年度年次大会

学会賞選考委員会委員長
投野由紀夫（東京外国語大学）

東支部活動予定

東支部では東京外国語大学グローバルCOEプログラムとの共催で以下の要領で大学院生対象の研究発表会を行います。テーマは学習者コーパスで、研究途中の中間発表でもかまいません。また対象言語も英語だけでなく他の外国語（日本語含む）でもかまいません。会議には、国際学習者コーパスプロジェクトICCIに参加している世界7地域の研究者が集い、院生の発表を聞いてコメントをしてくれます。大変有意義な会になると思いますので、奮ってご参加ください。

Postgraduate Conference on Learner Corpora

日時：2011年3月14日（月）10:00-18:00

場所：東京外国語大学

発表言語：英語

アブストラクト申し込み〆切：

2010年12月24日（金）

採択通知：2011年1月20日（木）

詳細は下記サイト参照：

<http://cblle.tufs.ac.jp/index.php?id=305>

東支部支部長

投野由紀夫（東京外国語大学）

今後の大会日程と開催校

第37回大会	2011年10月1日（土）、2日（日）	京都外国語大学
第38回大会	2012年10月	開催校未定
第39回大会	2013年10月	東北大学



事務局から

◇会費納入のお願い

2010 年度会費（一般 5,000 円，学生 3,000 円）を，日本郵便にある払込取扱票を使ってお納めいただきますよう，ご協力をお願いいたします〔振替口座：00940-5-250586〕。日本郵便発行の受領証をもって領収書に代えさせていただきますので，ご了承ください。別途領収書が必要な方は，80 円切手を同封の上，石川保茂（〒615-8558 京都市右京区西院笠目町 6 京都外国語大学）までお申し出ください。払込取扱票の通信欄によるお申し出はご遠慮ください。

過年度会費未納の方は，2010 年度分と併せてお納めください。会誌『英語コーパス研究』第 18 号は 2010 年度の会費を納入していただいた方のみ，送付いたします。また，2 年続けて会費未納の場合，*Newsletter* などの送付を中止させていただきます。

住所，所属などに変更や異動のある方は，必ず払込取扱票の通信欄にお書き添えください。

※会員の皆様には，日頃より会費の当該年度内納入のご協力をいただきまして，お礼申し上げます。会費を滞納されますと，退会時に滞納分をまとめてお支払いいただくといった事態にもなりかねません。会員の皆様におかれましては，円滑な学会運営のためにご協力いただけましたら幸いです。なお，退会を希望される場合は，少なくとも 4 月 30 日までに事務局までお知らせくださいますようお願い申し上げます。

◇メーリングリストについて

英語コーパス学会ではメーリングリストを使って会員の皆様の様々な情報交換に役立てていただいているところですが，最近，宛先不明でエラーが返ってくる例も増えています。会員の皆様方には，メールアドレスに変更が生じた場合，速やかに事務局宛ご連絡いただけますようお願い申し上げます。

◇寄贈刊行物の紹介

磐崎弘貞 (2010) 『英語語彙指導の実践アイデア集：活動例からテスト作成まで』大修館書店。
加野まきみ (2010) 『コーパスを活用した認知言語学』大修館書店。

☆☆☆☆☆☆☆☆☆☆

FORUM

■ICTATLL2010 Kyoto Conference 開催報告

石川慎一郎（神戸大学）

iskwshin@gmail.com

1. ICTATLL とは？

ICTATLL とは，ICT in Analysis, Teaching and Learning of Language の略で，言語研究・言語教育における ICT 活用について研究する学術団体です。年会費などを伴う定常的会員を持たないカンファレンスベースの学会で，本部はイギリスにあり，毎年，世界の各地で学会を開いています。決して大きな規模ではありませんが，これまでにイギリス・ギリシア・スリランカ・日本他などで，主として現地の研究者を集めて開催を重ねてきました。

2. ICTATLL2010 の概要

2010 年の第 6 回大会は日本で開催することになり，筆者の研究室がホストを引き受けました。大会テーマは Corpus, ICT, and Language Education，会場は京都キャンパスプラザ，会期は 2010 年 9 月 21 日から 23 日の 3 日間です。日本・中国・香港・台湾・タイ・韓国・イギリス・アメリカ・スペインから約 50 名の参加者があり，審査を経て採用された 27 件の発表（2 本の基調講演を含む）がありました。基調講演は東京外語大の投野由紀夫先生とストラスクライド大学のジョージ・ウィアー先生に行っていただきました。

大会後，独立した Scientific Committee による審査を経て選ばれた論文をまとめ，イギリスのストラスクライド大学出版局より，Weir, G., & Ishikawa, S. (Eds.), *Corpus, ICT, and language education* [246 pages] が刊行されました。人文系・教育系・工学系という学問の垣根を越え，コーパス・ICT・言語教育にかかわるすぐれた論文が多く集まり，学会の記録を形として残せたのは大変良かったと考えています。

なお，英語コーパス学会の先生方には，Scientific Committee 委員として，また，講演者・発表者・司会者・投稿者として，大変お世話になりました。この場を借りて厚く御礼申し上げます。

3. 発表論文の概要

以下、おおまかな分野にわけて、発表論文の概要を簡単に紹介しておきます。まず、「言語教育とコーパス」関連では、Yukio TONO 氏が学習者コーパス研究の最新の状況について、Shin'ichiro ISHIKAWA が開発中の学習者コーパスを用いた母語話者・非母語話者の助動詞使用について、Hui JIN 氏が学習者コーパスの中の 1 人称使用について、John MILTON 氏がコーパス研究の成果を言語教育に生かすシステム開発について、Kiyomi CHUJO 氏が統計手法を用いた分野特徴語の抽出法についてそれぞれ報告しました。

「言語研究とコーパス」関連では、Serene S H NG 氏がコーパスを用いたシンガポールマンドリンにおける語彙の揺れについて、Satoshi INOUE 氏が BNC に見る知覚動詞について、Yuka Ishikawa 氏が各種のレジスター下での性特定職種名称の出現について、Juan M. HUERTA 氏が数学教授における発話データの談話特性について、George R S WEIR 氏がコーパスから得られた量的言語指標を楽曲の構成単位に変換してテキストを楽曲化するアプローチの可能性について、Makoto SHIMIZU 氏が科学系の学術論文に出現したコロケーションについて、Hirosada IWASAKI 氏が 3 種類の映画スクリプトの語彙構成の特性についてそれぞれ考察を行いました。「言語教育と評価」関連では、Huiying YANG 氏が中国における英語教授の現状と課題について、Jichun PAN 氏が英語口頭発表におけるエラーの扱いについて、Toshihiko UEMURA 氏が英語における初年次教育の成果について、Praneet CHURAM 氏がタイの大学学生の英語談話力の発達について、Hitokazu MATSUSHITA 氏が新しいテスト理論に基づく効果的な口頭能力テストの開発について、Mireia FARRÚS 氏が連続的な習熟度評価テストの自動化システムの開発について、それぞれ分析と考察を行いました。

最後に、「ICT を用いた言語研究・教育」関係では、Chiou-hui CHOU 氏が EFL 教育における WebQuest の効果について、Kamollada RAKSASAB 氏が CALL を用いたイディオム習得について、Tomoko NAKAMURA 氏が初学者向けのチュートリアル教材の開発について、Katsunori KOTANI 氏が読解速度テストの開発について、Yoshinori MIYAZAKI 氏が技術英作文の支援システムについて、

Hiromi OYAMA 氏他が日本語の case particle の誤り自動検出システムについてそれぞれ報告と検証を行いました。

発表後には、同じ分野の、あるいはまた、異なる分野の参加者から多様な質問が出され、充実したディスカッションが展開されました。

4. 国際学会をホストして：体験の継承

報告者は、これまでに各地の国際学会に参加した経験はあるものの、国際学会の開催を責任者の立場で引き受けるのは初めての体験で、約半年にわたる準備の過程で多くのことを学びました。

たとえば、内外の関係団体への後援申請、ウェブサイト作成、海外での宣伝・広報、内外の発表者・寄稿者との連絡、海外参加者からの参加費の受領に伴う口座の開設や海外銀行との折衝、海外への経費送金、海外の大学出版局との交渉、ビザの発給にかかわる大使館・領事館とのやりとり、身元保証、所属先大学との折衝、宗教的な禁忌食の調査と対応したメニューの開発依頼などなど、学会規模が小さいためにほとんどの作業を一人でやることになり、苦勞と同時に大変に勉強になった半年間でした（半年の間に、それまでの 10 年分と同じぐらいの量の英語メールを書いた気がします…）。

準備作業をしていて実感したのは、国内の各地で各種の国際学会が開かれているにも関わらず、こうした開催ノウハウがほとんど蓄積・公開されていないことです。工学系や医学系の国際学会であれば専門業者が準備作業の一切を相応の経費で受注することが一般的だということですが、人文系の学会の場合、規模の問題もあって、やはり、責任者となった教員個人ないし教員グループが準備の仕事の多くを担うことになると思われます。ところが、そうした立場の人に必要となる情報はきわめて得にくいのが現状です。

報告者の場合も、身元がはっきりしない海外参加者から次々に身元保証や帰国旅費保証を求められ、さらには亡命相談(?)まで届くに至り、学会の終了まで非常に不安な思いをしました。大学教員が学会開催に際して行う参加者の身元保証について、その効力や責任の範囲をわかりやすく説明してくれる本やサイトなどがあればよかったです。なかなかそういうものはなく、「身元保証、責任」などのキーワードでネットを検索すると、「連鎖倒産」「

家離散」などの言葉ばかりがずらりと並び、不安は解消されるどころかいたずらに増幅されるような有様でした。

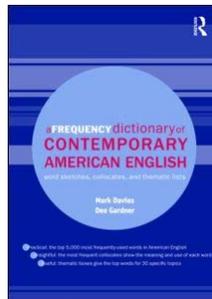
おかげさまでなんとか無事に学会を終えることができましたので、可能であれば、今後、同様の仕事をなさる方のために、「国際学会準備マニュアル」のようなものを作ってウェブなどで公開できればと思っています。その節はまた、過去に同様のご苦勞をなさった会員諸氏からお知恵を拝借できれば幸いです。

■新刊紹介

中條清美 (日本大学)

chujou.kiyomi@nihon-u.ac.jp

Davies, M. and D. Gardner (2010)
A Frequency Dictionary of Contemporary American English: Word Sketches, Collocates, and Thematic Lists. London and New York: Routledge. 368 pages.



£19.99 (paperback)

ISBN: 978-0-415-49063-4

本書は、The Corpus of Contemporary American English (COCA) の約 4 億語 (1990–2009) の言語資料を分析した語彙の使用頻度調査より作成された 5,000 語の辞書／語彙リストです。いわゆる語句の意味を調べる「辞書」ではなく、大規模コーパスの研究である Francis & Kučera (1982) のような「語彙リスト」に頻度数など数種類の情報がコンパクトに付加されたものです。

本書は 3 部構成で、5,000 語の Frequency index (pp. 9–281), Alphabetical index (pp. 282–316), Part of speech index (pp. 317–351) からなっています。メインとなる Frequency index には、各見出し語の「rank, レマ, 品詞, collocate, raw frequency, dispersion, ジャンル」が記されています。

本書の作成目的は、①最新のデータに基づいて、②最頻出の 5,000 語を選定し、③それらがどのような語と一緒に出現するかを明らかにすることでした。本書に基づいている COCA は spoken (79 mil.), fiction (76 mil.), popular magazines (81 mil.), newspapers (76 mil.), academic journals (76 mil.) の 5 分野のサブ

コーパス、計 150,000 texts より構成されています。

本書の特徴は、①現代のアメリカ英語に基づいていること、②書き言葉と話し言葉の均衡コーパスであること (話し言葉は 1/5)、③見出し語には品詞別に 20~30 の collocates (隣接する語) を提示して意味と用法の情報を提供していること、④トピック別に特徴度の高い語をリストした 31 の ‘thematic box’ コラムを付けているところにあります。

Introduction には、大規模コーパスの頻度分析に基づく 5,000 語の語彙リストの作成手順がわかりやすくまとめられています。まず、CLAWS tagger (Version 7) を使用してタグ付とレマ化を行いました。次に、コーパスを 4 mil. × 100 section に分け、サブコーパス間の分布の尺度として Juillard’s “D dispersion index” を用いて、最終的に、score = frequency × dispersion を求め、score 上位 5,000 語を見出し語としました。注には同綴異義語や固有名詞等の問題点の処理方法とその限界について書かれています。

見出し語の一例として bag (p. 64) を示します。bag の場合、ランクは 1045 位で、品詞は名詞です。bag と一緒に出現する collocates には、出現頻度順に、形容詞は brown, full, . . . , 名詞は plastic, air, . . . , 動詞は carry, put, . . . 等があります。Plastic や air の右側のピリオド (.) は「bag の出現位置」を示し、bag は plastic bag や air bag として現れることを示しています。Collocates の抽出方法は、たとえば、中心語 (node word) が動詞の場合、中心語の前後 4 語の幅 (context or span) において、「名詞」と「その他の品詞」の collocates を調査し、また、中心語が名詞の場合は「形容詞」、「その他の名詞」、「動詞」の collocates を調査した後に、それらの調査結果を 4 名の英語母語話者が評価しました。なお、the, of, to のような語については統計指標 MI (自己相互情報量) を用いて collocates を抽出したということです。

1045 bag *n*

adj brown, full, heavy, mixed, overnight, punching, zipper-lock, athletic **noun** plastic., air., sleeping., paper., shopping., garbage., duffel. **verb** carry., put, pack, hold, fill, pull, open, grab., drop, pick.
36465 | 0.92 F

見出し語の collocates の後には、bag の出現頻度

36465 と、bag のサブコーパスにおける分布の広がり度を表す dispersion 0.92 (1.00 はすべてのサブコーパスに現れる) が付けられています。最後の F は Fiction を示しています。特定のジャンル (この場合、Fiction) において出現頻度率が他の 4 ジャンル (この場合、Spoken, Magazines, Newspapers, Academic) と 2 倍以上の差があった場合には、上記の bag の例のように、最後に頻出ジャンルのイニシャルが付けられています。

次に、31 個の thematic box の一例として、21 番目 (p. 150) に付けられた、「米語新語リスト」の一部を示します。E-mail や terrorism 等の単語の後の数字は出現頻度を表しています。これらの語は 90 年代と比較して 2000 年代に「3 倍以上多く現れた」新語のリストであり、政治的、科学・技術的な社会の変化が語の使用頻度の変遷にも反映されていることがわかります。Thematic box には他にも、animals, body, clothing, food 等の身近なトピックの語彙リスト、COCA の各分野の特徴語や、各品詞の接辞リスト等興味深い語彙が含まれています。

21. New words in American English

[nouns] e-mail 14326, terrorism 10360, terrorist 8889, affiliation 8713, adolescent 7212, homeland 4157, website 3909, Sunni 3637, wireless 3492, prep 3362, Taliban 3006, insurgent 2433, globalization 2354, SUV 2068, RPG 1970, anthrax 1954, steroid 1898, genome 1867, blog 1765, detainee 1733, militant 1610, ethanol 1601, insurgency 1561, yoga 1533, recount 1491, cleric 1466, coping 1380, tsunami 1310, cellphone 1149

2009 年 10 月にカナダのアルバータ大学で開催された American Association for Corpus Linguistics (AACL) における研究発表において、本書の著者である Davies と Gardner が本書の作成過程と本書の特徴を紹介していました。とりわけ、上記に述べた collocates の同綴異義語を見分けるために大量のコンテキストを読まなければならなかった苦労を Gardner が強調していたのが印象に残っています。頻度分析は語形に依存する form-based (lemma) ですから、たとえば lead (導く) と lead (鉛) のような同綴異義語は一つの見出し語のもとに記載されてしまうこととなります。そこで次善の策として、collocates の情報を付けることによって、ある程度それぞれの意味と用法を示すことができると彼らは説明していました。

Introduction の参考文献に挙げられている多くは、単語の使用頻度調査で著名な Thronkike & Lorge (1944), Rinsland (1945), West (1953) 等や、近代の代表的な大規模コーパスである Carroll et al. (1971), Francis & Kučera (1982), Johansson & Hofland (1989) 等の研究と、それに英語教育分野において頻繁に言及される Nation (2001), Coxhead (2000) 等で占められていて、本書がこれらの先人の広範な研究成果を念頭に置きながら編纂されたことがうかがわれます。実際、COCA の 5,000 語は、伝統的に英語教育の基本語彙選定において用いられてきた「より多く、より広く出現する語は重要である」という明解な基準に基づいて選定されたものであり、「基本語彙」の資料として参照可能な 5,000 語リストが選定されると期待されます。

本書掲載の 5,000 語等の電子データ ('similar data') はダウンロード可能です。ただし 9 月 1 日時点では、ダウンロード版の 5,000 語リスト (<http://www.wordfrequency.info/>) には空欄等が数カ所含まれていました。筆者が本書の第 3 部 Grammar Index の 5,000 語をスキャンした 5,000 語を、ダウンロード版 5,000 語と比較したところ、約 280 語の不一致が認められました。'similar data' (p. 7) と断られている理由は、そのような不一致があることを認めているためと思われます。本書の 5,000 語は利用目的によってはこうした不一致を考慮に入れておく必要があります。また、本書に基づく COCA の延べ語数は本書のカバー、前書き、そして Introduction の pp. 1-2 には 385 million words と記されていますが、Introduction の pp. 3-5 では 400 million words と記されています。Thematic box の個数はカバー等では 31 個、Introduction の p. 2 では 30 個、p. 7 では 31 個と記されていることにも注意が必要と思われます。

本書の対象は学生や教師向けとされています。パーペラバック版は値段、サイズも手ごろで、見易く、collocates をはじめとする利用価値の高い情報がコンパクトに盛り込まれていますので、語法研究、教材作成の際の参考資料として利用できます。また、thematic box などは読み物としても興味深く、*A Frequency Dictionary of Contemporary American English* はおすすめしたい一冊です。



■KWIC Concordance for Windows 新版の公開とプログラミングのすすめ

塚本 聡 (日本大学)

tukamoto@chs.nihon-u.ac.jp

KWIC Concordance for Windows の変更点

10年以上も前に KWIC Concordance for Windows を公開し、その後少しずつ機能の追加を継続してきた。当該ソフトは C++言語を用い、マイクロソフトの Visual Studio を利用し開発してきたが、Visual Studio のバージョン更新により、ある時点から過去のソースコードをコンパイルするとエラーが発生するようになった。過去のソースコードがそのまま利用できなくなったため、更新が滞るようになった。そのため、今回ソースコードを一新し、必要な機能のみに絞った新版を開発することとした。

今回の開発版から、C++による .Net Framework を使用した Windows Form を利用したスタイルとした。その結果、出力をセル形式のウィンドウに出力するようになり、キーワードなどの位置が固定化さ

れ、ソートやコピーなどがやり易くなった (図 1, 2)。

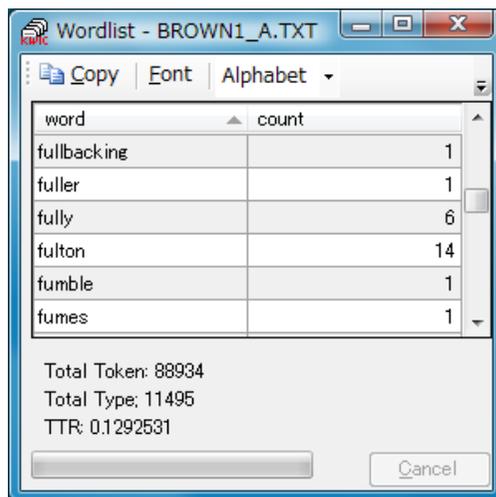


図 1 単語リスト



図 2 コンコーダンス

AntConc や WordSmith のような豊富な機能はなく、基本は単語リスト、コンコーダンス、コロケーションの 3 機能で、対応するコーパスは、テキスト、固定長、COCOA、SGML など旧版対応のなかから利用頻度の高いものに限定した。ただし、Brown, Frown などよく利用するコーパスはメニューから選択し、環境設定が可能となっている。

旧版にあった参照部の情報を取り出す機能は本版でも有効であるので (図 2)、他のソフトなどと連携すれば、参照情報を活用できるであろう。

旧版では、利用可能なコーパスや機能を順次追加していったため、同種の処理が各所に分散し、無駄の多い作りとなっていた。また変更の際には各所に修正を加えなければならなかった。その経験を生か

し、今回は基本部分と各コーパス固有の部分とをあらかじめ分離し、中核部分とコーパス固有の部分とを接続することを念頭に置きながら全体を作る方針とした。そのため、ある処理をするために、その処理だけを考えればもっと単純化できる部分でも、統一をとるために無駄な関数などを設定するような部分もある。

本バージョンから .Net Framework と呼ばれる仕組みを利用するよう変更したため、しばらくはアップデート可能と思われる。また、内部では文字列を Unicode で処理しているので、Unicode で保存したファイルであれば日本語も扱えるようになった（ただし、分かち書きをしていないと、語の区分ができない）。現段階では基本機能はほぼ完成しているが、処理中断のためのキャンセル機能が一部不完全である。今後はこの部分の改良に努めたい。また、ソフトウェアをご利用いただいた際に問題点があればお知らせいただければ幸いである。

プログラミングのすすめ

第 34 回大会のワークショップにおいて python を用いた処理が紹介されていたように、既存のソフト

ウェアでは扱いにくい処理を行う際には、研究者が独自にプログラミングをする必要がある。そのような用途に awk, perl などのスクリプトを利用される方も多いと思われる。しかし、C++や C#などの言語を利用する方は、多くないのが現状であろう。その要因に、「難しい」という印象をお持ちの方も多いと思われる。確かに、本格的に利用しようとする、多くの関数や規則を把握しておかなくてはならないのも事実である。しかし、コーパスや言語処理に関連する文字列の処理という限定された範囲内であれば、それほど多くの事柄を気にすることはない。ここでは、C#というプログラム言語の例として（厳密には .Net Framework に依存した例）、単語リスト作成のプログラムを示し、プログラミングは予想より簡単であり、さらなる利点があることを紹介したい。

下記の例は、“wordlist” という名称で、C#のコンソールアプリケーションを作成し、そのひな型に、①から⑧の次のカッコまでの部分を追記したものである。キーボードやファイルから文字列（文章など）を読み込み、単語に分割し、単語リストを作成する処理が中核部分となる。

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;

namespace wordlist
{
    class Program
    {
        static void Main(string[] args)
        {

SortedDictionary<string, int> words = new SortedDictionary<string, int>();           ①
string line;

Console.WriteLine("Enter one or more lines of text (press CTRL+Z to exit):");

while((line = Console.ReadLine())!=null)                                         ②
{
string [] split = line.Split(new Char [] { ' ', ',', '.', ':', '\t' });           ③
foreach (string s in split)                                                       ④
{
try
{
words.Add(s.ToLower(), 1);                                                       ⑤
```

```

}
catch (ArgumentException)
{
words[s.ToLower()]++;
}
}
}

foreach (KeyValuePair<string, int> kvp in words)
{
Console.WriteLine("Key = {0}, Value = {1}", kvp.Key, kvp.Value);
}
}
}
}

```

ここに記載した処理のみで、実際に単語リストを作成することができる。このうち、中核となる処理の概略は、以下のとおりである。

- ①perl の連想配列に相当する単語リストを記録する変数の宣言
- ②コンソールからの 1 行入力・ループ
- ③複数の単語を含む行を分割し、各語を文字列配列に代入
- ④配列内の各語が無くなるまでループ、各語を s という変数に代入
- ⑤変数 s にある語をキーに単語リストに生起数 1 として追加
- ⑥すでにリストにある場合は、リスト内の数値に 1 を加算
- ⑦単語リスト変数内の語がある限りループ、各語と生起数を kvp という変数に代入
- ⑧語と生起数をコンソールに出力

perl に比べると、変数の宣言や型の指定 (string や int) など手間のかかる部分もあるが、それほど大きな違いが無いことがお分かりいただけると思う。とりわけ、perl などの利点となる連想配列が、C#では SortedDictionary コレクションという仕組みを用い、同様の容易さで利用可能である。また、単語の切り

分けも、細かなプログラミングを必要とせず、1 行で行うこともできる。出力も全体の語数などは考慮せず、処理することができる。

さらに、プログラミングに関する情報がオンラインなどで豊富に提供されている。今回紹介したプログラムは、マイクロソフトの MSDN ライブラリ (<http://msdn.microsoft.com/ja-jp/library/default.aspx>) から、“SortedDictionary クラス”を検索し、その定義・説明のページに例示されていたものをコピー&ペーストし、ひな形として利用したものである。ここでは、単語リストとして見出し語とその頻度をデータとして扱うので、SortedDictionary を<string, int>というデータ型に変更し使用した。同様に、WriteLine や Split も、ヘルプにある使用方法、使用例に多少の変更を加え利用している (Split についてはそのまま使用しているので、単語の区切り定義としては不十分である)。

プログラミングで一番手間のかかることは、意図したとおりにプログラムが作動せず、その修正を行うことである。その原因は、計画通りに変数に数値や文字列が代入されていないことが多い。プログラムを修正することをデバッグと呼ぶが、今回使用した Microsoft Visual Studio のような統合環境では、充実したデバッグ機能が特徴である。

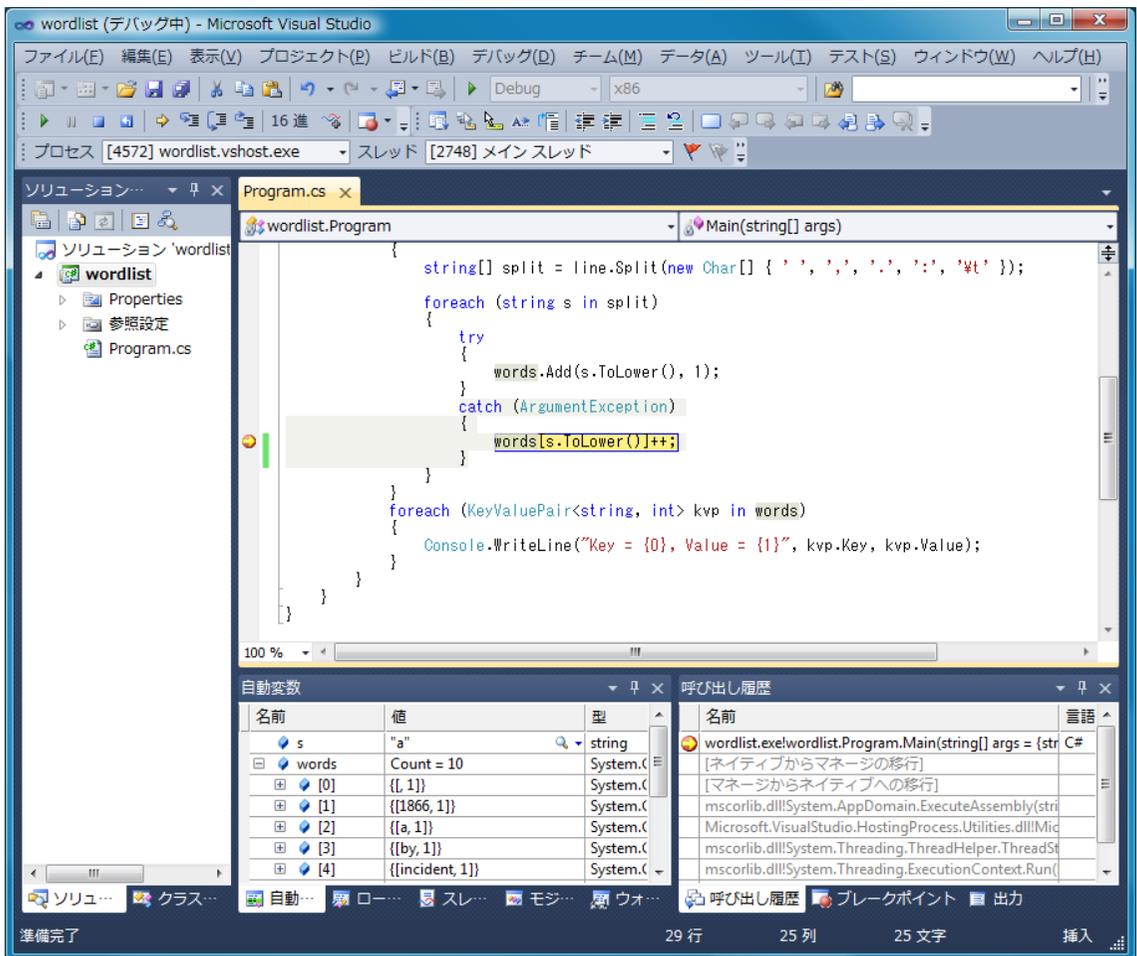


図3 デバッグ状況

図3は、“a”という単語が2回目に出現した際のデバッグ状況を示した画面である。⑥にあたる部分で一時停止するよう設定したので（画面上、○および⇒の印がある）、その段階で進行が停止している。画面下部の「自動変数」欄を見ると、現在使われている変数が表示されている。“s”はプログラム内で、単語を取り出す入れ物として使っている変数で、“a”が代入されている。“words”はコレクションと呼ばれる配列で、すでにいくつかの語（“1866”、“a”、“by”など）がソートされた状態で代入されていることが示されている。“a”がすでに words 配列にあるので、意図したとおり⑥の処理が行われようとしていることが確認できる。マウスでボタンをクリックすることにより1行ずつ処理を進めることができるので、⑥の処理を終えると words 配列中の [2] の部分が {[a, 2]} のように変化し、生起数が加算されていることも確認できる。このように、意図

したとおり、文字が分割されているか、変数に代入されているか、変数の値が加算されているかを確認することが極めて容易である。統合環境でのデバッグ作業が容易であることは、perlなどのスクリプトに対する大きな利点となる。

この統合環境である Visual Studio では、IntelliSense と呼ばれる関数の補完機能もあり（Excel の関数補完機能と同様のもの）、一部を入力すると残りの部分が自動的に補われる。さらに調べたいコマンドにカーソルを置いて F1 キーを押すとヘルプへリンクするようになっている。おぼろげな記憶でも、順次情報を入手、確認しながらプログラミングを進めることができる。

さらに、多くの方が利用している Windows 用には、Visual Studio Express として無料のソフトウェアが提供されている（現在は Visual Studio 2010 Express）。perl なども無料で使用可能であるが、統

合環境までは提供されていない。他の統合環境などと合わせて使用すればデバッグなども容易に可能となろうが、Visual Studio は、初めからすべてがそろった形で無料提供されている。

ここでは、C#による単語リストのプログラム例を示した。冗長と思われるが例となるソースコードをすべて掲載した。その目的は、処理自体はかなり単純なコマンドで行えること、プログラムやデバッグが容易であること、開発環境が無料で利用可能で

あることを示すことにあった。とりわけ、perl などのスクリプトの経験者であれば、ループや分岐などの処理はよく似た記述方法をとっており、ほとんど違和感を感じないであろうし、その特徴の一つである連想配列の機能は、ほぼ同様の容易さで C#で利用可能である。perl から C#や C++へは、ほんのわずかの移行であることを感じていただけたものと思う。多くの方にプログラミングに取り組んでいただきたいと思っている。