英語コーパス学会
Newsletter No. 83
Dec. 28, 2017

■会長: 投野　由紀夫
■事務局: 〒157-8511 東京都世田谷区成城6-1-20　成城大学社会イノベーション学部　石井康毅研究室気付
■郵便振替口座: 00930-3-195373 英語コーパス学会）
■URL: http://jaecs.com/ ■e-mail: jaecs.hq@gmail.com ■twitter: @JAECS2012

JAECS
Japan Association for English Corpus Studies

＜第 43 回大会報告＞

■概要

英語コーパス学会第 43 回大会は，2017 年 9 月 30 日（土）と 10 月 1 日（日）の 2 日間にわたり関西学院大学（西宮上ケ原キャンパス）にて開催されました。第 43 回大会では，石川慎一郎先生（神戸大学）の招待講演，長谷部陽一郎先生（同志社大学），小林雄一郎先生（日本大学）によるワークショップ，野口ジュディー先生（神戸学院大学名誉教授）の司会進行によるシンポジウム，さらに 21 件の研究発表が行われるなど，研究発表とワークショップの数が前年度より増え，充実した大会となりました。

まず，大会初日の午前中には，長谷部陽一郎先生によるワークショップ「TCSE を用いた TED Talks の全文検索と英語教育への応用」が開かれ，長谷部先生がご自身で開発した TCSE の解説がなされました。

1 日目の大会は，まず投野由紀夫会長（東京外国語大学）による開会の挨拶に続き，開催校である関西学院大学副学長の小菅正伸先生に挨拶のお言葉をいただきました。

次に，内田充美先生（関西学院大学）の司会のもと総会が行われました。始めに，石井康毅事務局長（成城大学）より春・秋の理事会での決定事項のうち，人事・規則についての報告がなされました。次に，東支部の解散と研究会（SIG）の設置，及び理事年会費の廃止に関する会則の改定が諮られ，承認されました。続いて，投野由紀夫会長から新設の研究会についての紹介がなされました。最後に，会計の宇佐美裕子先生（東海大学）より，2016 年度会計報告及び 2017 年度予算案が示され，いずれも承認されました。

総会に続き，学会賞選考委員長の西村秀夫先生（三重大学）から学会賞および奨励賞については応募がなかったことが報告されました。

第 1 日目は午前中に長谷部陽一郎先生のワークショップが行われました。午後の研究発表は 3 室でのパラレルセッションとなりました。前半（13:30‐15:45）は，第 1 室 佐竹由帆先生（駿河台大学），第 2 室 藤原康弘先生（名城大学），第 3 室 後藤一章先生（摂南大学）の司会のもと，計 12 件の研究発表が行われました。後半（16:05‐17:45）は，第 4 室 石川有香先生（名古屋工業大学），第 5 室 大谷直輝先生（東京外国語大学），第 6 室 森下裕三先生（環太平洋大学）の司会のもと，計 9 件の研究発表が行われました。

2 日目の午前は小林雄一郎先生（日本大学）によるワークショップ「機械学習を用いたコーパス分析入門」が行われました。その後，石川慎一郎先生の講演「A Frontier in Learner Corpus Studies: For Better Understanding of L2 Learners」が行われました。大会の最後に，野口ジュディー先生の司会進行でシンポジウム「話し言葉コーパスの構築と利用」が行われました。石川慎一郎先生，迫田久美子先生（広島大学・国立国語研究所），野口ジュディー先生，長谷部陽一郎先生の 4 名の先生方から話し言葉コーパスの構築に関するご発表をいただきました。また，発表後には質疑応答の時間も設けられました。

第 43 回大会は 2 日間を通して，138 名の参加者がありました。多くの皆様にご来場いただき，質量ともに充実した学会を開くことができました。ありがとうございました。

■研究発表セッションの概要

*印のついた発表を除き，研究発表の概要は各発表者が作成した要旨を司会者が編集したものです。質疑応答についての記述は司会者が執筆しました。

第１日第１セッション
［司会・報告］佐竹由帆（駿河台大学）

「Charles Dickens の *The Mystery of Edwin Drood* と Thomas Power James によるその続編の文体類似性評価」

後藤克己（中部大学大学院生）

Thomas Power James（以下，T. P. James）は *The Mystery of Edwin Drood*（以下，*ED*）の続編を発表し，自身で“By the Spirit Pen of Charles Dickens, through a Medium.”とアピールしている。本発表では *ED* と続編との語彙使用面での文体類似性について数量的に比較した結果を報告した。

ED とその続編に，*Our Mutual Friend*（以下，*OMF*）を加えたコーパスについて，章ごとに地の文に生起する語彙のレマでの頻度を抽出したデータで対応分析した結果，*ED* と *OMF* の各章の散布は重なって一つのクラスターを形成し，続編のそれは明らかに別クラスターを形成していることが明らかになった。また，*ED*・続編・*OMF* のフルテキストのコーパスによる比較で，Mahlberg (2013)で Dickens 作品に特徴的とされている 5 語クラスター（Body Part, As If）の生起頻度は *ED*／*OMF* と続編とで大差があること，さらに類義語 as if / as though については *ED*／*OMF* と続編とで嗜好が逆になっていることが分かった。これらの結果から続編の *ED* との文体類似性は低く，“By the Spirit Pen of Charles Dickens, through a Medium. ”との T. P. James のアピールは疑わしいことを指摘した。

質疑では，発話部の語彙の比較についての質問があり，対応分析での比較は行っていないものの，JACET8000 による語彙の難易度分析ツールを用いて，*ED* と続編について地の文と発話部の語彙の豊かさ（Index/Token 比率）を比較したところ，*ED* とは異なり続編では，地の文と発話部とで差が殆どなかった旨回答された。

なお，発表後，発話部を除外して地の文を抽出する方法について，発表者から詳細な補足説明があった。司会者からの，発話部にも作者の文体は反映するのではないか，という質問に対しては，本研究の語彙頻度比較では扱わなかったが今後考慮する旨を発表者は回答した。

「TF-IDF を用いた Alice Bradley Sheldon の文体分析」

木村美紀（明治大学大学院）

本発表では，男性であると思われていた作家である Alice Sheldon の文体に関して，TF-IDF と高頻度語彙の unigram，bigram，trigram という 6 種類の指標を使用しながら計量文体分析を行った。分析には先行研究で使用した Alice Bradley Sheldon 全 72 作品を収録したコーパス（延べ 865,802 語）と Ernest Hemingway 69 作品を収録したコーパス（延べ 271,475 語）に加えて，Alice Sheldon と同時代・同ジャンルで活躍していた Ursula K. Le Guin の 45 テキストを含むコーパス（延べ 589,481 語），Octavia, E. Butler の 93 テキストを含むコーパス（延べ 867,396 語），Arthur C. Clarke の 104 テキストを含むコーパス（延べ 467,983 語），Theodore Sturgeon の 70 テキストを含むコーパス（延べ 475,704 語）を使用した。

サポートベクターマシンとランダムフォレストという 2 種類の統計手法を使用しながら分析を行った結果，TF-IDF よりも高頻度語彙の方がよい分類結果を得られるということが判明した。また，この結果から Alice Sheldon の文体が，男性作家群に類似しているのか女性作家群に類似しているのか検証した結果，Alice Sheldon の 72 作品中 13 作品が誤判別されていた。具体的には 13 作品中 6 作品が Theodore Sturgeon に，5 作品が Arthur C. Clarke に誤分類されています。また，1 作品ずつ Ernest Hemingway と Ursula K. Le Guin に誤分類されている。誤分類されたものの 92%が男性作家作品群へと判別されていることから，この 6 人の作家作品群という小規模なデータセットを使用した分析では，Alice Sheldon の文体は女性作家よりも男性作家の文体に近いと結論付けられる。この結果は，Silverberg (1975)や小谷 (1994)などが主張している「Alice Sheldon の男性的な文体」ということを計量的に再確認できたと考えている。今後，どのような言語特徴が著者判別指標となりうるのか詳細に検討する必要が存在する。

発表後，TF-IDF の実装方法と指標の選択方法について発表者から詳細な補足説明があり，フロアからの方法論に関する質問に発表者は回答した。Sheldon の文体は男性的であるが Hemingway と類似しているとは言えないという結論は，計量的分析を経てはじめて導き出されるものであり，従来の主観的文体論にはできない，文学批評への有意義な貢献であると言えるだろう。

「機械学習アプローチによる小説テクストの計量的分析：アーサー・コナン・ドイルの作品から」

黒田絢香（大阪大学大学院）

本発表では，機械学習の手法を用いて Arthur Conan Doyle の作品を分析する研究の結果について，特にトピックモデルと従来手法との比較を中心に報告された。

テクスト間で使用頻度に統計的有意差のある語を抽出する特徴語抽出の研究は数多く行われてきたが（高見, 2004），近年では機械学習アルゴリズムの一つであるトピックモデルが注目されている。

まずトピックモデルのアルゴリズムの仕組みを提示したのち，文学分析に応用した先行研究が挙げられ（Jockers and Mimno, 2013; Tabata, 2017），モデリングの前処理手法や，出力結果の可視化手法について論じられた。

次に，実際に Doyle の作品群を対象としてモデリングを行い，推理小説 7 作品と歴史小説 9 作品のそれぞれに特徴的に多く出現するトピックがどのようなものかを幾つか挙げた。

例えば，推理小説群では case, evidence, police, found などの語を中心としたトピックが高確率で出現しており，このトピックは“Criminal Investigation”とラベリングできる。

この結果と，従来手法である対数尤度比検定などで得られた結果を比較し，トピックモデルを用いた場合では関連性のある語が比較的低頻度の語も含めて効率的にまとめられていることが示された。

また，ネットワークグラフを用いて複数のトピック間の関係からより探索的に特徴を抽出していくアプローチを提案した。

質疑では，文学研究に機械学習を用いることの意義や，得られた知見からどのように議論を展開するかについて質問があり，文学分析における位置付けを明確にする課題が明らかとなった。

なお，発表後，分析手法と結果について，発表者から発表時間内では扱わなかったネットワークグラフも提示しながら，詳細な分析結果の補足説明があった。

「Agatha Christie 作品の修辞的特徴に関する分析」

土村成美（大阪大学大学院生）

本発表では，Agatha Christie 作品の文体的特徴について，修辞的項目を指標とした計量的分析を行なった。文学作品の計量的な分析では，語や品詞の使用率を指標とすることが多く，修辞的項目を指標とした分析はそれほど多くない。そのため Agatha Christie 作品と比較対象とした Dorothy Sayers の作品のそれぞれに修辞的アノテーションを行い，それを元に機械学習を用いた分析・考察を行なった。

分析対象は Christie 221 作品（5,071,282 語），Sayers 55 作品（1,375,645 語）から，各作家につき 50 作品ずつ無作為に選定した。

本研究では修辞的アノテーションを行うにあたり，DocuScope を使用した。DocuScope は Kaufer & Butler (1996)におけるレトリック理論と Kaufer & Butler (2000)における言語表象理論を基礎として構築されたツールであり，言語表現を 101 の Language Action Types (LATs)に分類し，タグ付けを行う。このタグ付けを通して得られた LATs の頻度情報を変数として，機械学習の一種の Random Forests を用いた分類を行なった。Random Forests の分類において寄与率の高い項目に関して検討を行った。

Sayers 作品と比較して Christie が最も過剰使用しているのは，一人称代名詞と動詞や前置詞の組み合わせで一人称としての意識を表す Self-Disclosure (I think, I feel, for me など)であった。Sayers と比較すると Christie 作品では会話文だけでなく，地の文でも一定数使用されていることが確認された。Christie 作品では作中の登場人物が語り手を務める作品も多く，両作家の地の文での語り手の違いが反映されている可能性があると考えられる。また，Christie は熱中や傾倒を表す Intensity (very, indeed, I do など)も多く使用していることが確認された。これについては同一作品における同一語の繰り返しが顕著に見られ，Christie の語彙多様性が晩年になるに従って低下したことを反映していると考えられる。

発表後，クリスティの語彙多様性が晩年になるに従い低下したことと彼女がアルツハイマー病であった可能性との関連について，フロアから，同じ語の繰り返しが作品上必要な場合もあるのではないか，との指摘があった。計量的手法が文学解釈の領域をどのように広げていけるか，を問うコメントもあった。

第１日第２セッション
［司会・報告］藤原康弘（名城大学）

Investigating the Effect of Data-Driven Learning on Lexico-grammatical Proficiency in an Extensive Reading Program

Gregory Hadley（新潟大学）
ハドリー浩美（新潟大学）

This research reports on an ongoing project investigating the use of what they call ‘soft’ version of data-driven learning (DDL) (Hadley & Charles, 2017), as a means of stimulating greater lexicogrammatical knowledge and reading speed among lower proficiency learners in an extensive reading program. According to them, it is the first study to investigate the effects of DDL on receptive learning.

From April 2015 to July 2017, students from six extensive reading classes were chosen for this study. For 16 weekly 90-minute sessions, an experimental group (21 students) used DDL materials created from a corpus developed from the Oxford Bookworms Graded Readers, which contained 186 books from all seven levels with a total of 1,715,160 tokens (17,670 word types). The control group (28 students) had no DDL input. Students read a minimum of 200,000 words during the course. Students not reaching the 200,000 word threshold were removed from this study.

A C-test (Klein-Braley & Raatz, 1984) and a speed reading test by Quinn, Nation, & Millett (2007) were selected to collect the data. A pre-test post-test design was chosen, and t tests were used. Pre-test analysis found that the experimental group was statistically lower than the control group. Post-test scores found that, within both groups, the learners improved significantly, but the experimental group improved to the point that they entered the same statistical bands as the control group. Post-test findings also indicate that students using the DDL materials were reading more and faster than the control group. With the results above, they conclude that the creation of DDL materials stimulated deeper levels of learning. Materials development required a significant amount of time, and despite limitations within using a flexible version of DDL, they believe that Data-Driven Learning can improve receptive learning and lexicogrammatical proficiency over that of extensive reading alone.

「日本人英語学習者の関係代名詞の回避：CEFR レベルを用いた検証」

高橋有加（東京外国語大学大学院生）

本研究の目的は，日本人英語学習者 93 名による①通常の英作文と，②関係詞を使うように指示のある英作文の，2 つのタスク内での関係詞の使用頻度とエラー率を比較し，知識はあっても使用が回避される傾向が CEFR レベル別にどの程度あるのかを明らかにすることである。

関係詞は日本人英語学習者にとって難しいものであり，多くの学習者は回避する傾向にあるといわれている（Shachter, 1974）。また，ヨーロッパ言語共通参照枠（Common European Framework of Reference: CEFR）の 6 つのレベルを同定する言語特性である基準特性（criterial features: Hawkins & Filipović, 2012）に関して，関係詞の使用が有効であることが指摘されている（Hawkins, 2009）。

しかし，自発的な産出が求められるパフォーマンステストなどでは，知識があっても実際に産出されないと評価することができないため，自然な産出及び意識的な産出における頻度とエラーについて比較した。研究設問は①関係詞の回避または不使用の現象が実際にどのくらいあるのか，②意図的に関係詞を使用するようにした場合にはエラー率は増加するのか，の 2 点である。

調査の結果，関係詞を使うように指示のあるタスクでは全レベルで関係詞の頻度が大幅に増加した。このことから，自然な産出と学習者が潜在的に持つ知識との間には差があることが確認された。また，関係詞を使うように指示のある英作文では A2 レベルでのエラーが多く観察されたことから，A2 レベルにおける回避が顕著である可能性が示唆された。さらに，上級学習者においては，関係詞を使用するよう指示があっても，不自然だと判断した場合は使用しない傾向があることが判明した。

質疑応答においては，個々の学習者の関係詞の使用回数の偏り，所有格の階層性等の考慮について話し合われた。

「英語教材としてのアニメ分析」

寺島芙由（東京外国語大学大学院生）

本発表では，アニメを英語学習に活用することの可能性について，語彙分析の結果に基づき，報告がなされた。アニメは学習者の興味関心を高める有効な教材となり得ると期待される。しかしアニメを利用した英語学習の研究は

ほとんど行われていない。そこで，英語に翻訳された日本のアニメ 4 作品と，アメリカで制作された映画 1 本を，以下の 3 つの観点から比較分析を行った：①アニメの視聴が学習に適しているか；②アニメと映画にそれぞれどのような特徴があるか；③アニメによる自主的な学習のためにどのような援助を行うのが効果的か。

その結果，①については，BNC spoken corpus の高頻度語や高頻度 N-gram がアニメにも高頻度で現れたことから，アニメは良質なインプットであると氏は主張する。

②については，氏は，ほぼ同時間の長さのアニメと映画のテキストを比較すると，アニメは総単語数が多く，かつ同じ語の繰り返しが多い傾向にあると述べた。

最後に③については，氏は固有名詞やアニメ特有の用語，また，アニメの内容に基づいてとりわけ多くみられるコロケーション（例：wish for）があることを示した。これらの単語や表現を事前に提示することで，学習者の負担を減らし，学習を促すような援助方法が考えられると主張した。

質疑応答時には，「アニメ・コーパス」の代表性について質問があった。比較対象の映画のスクリプトや，BNC の妥当性についても質問がなされた。英語学習教材としてのアニメの活用は興味深い分野であるとのコメントもなされた。

「小中連携に向けた英語授業コーパスデータ構築とインタラクション分析の試み」

片桐徳昭（北海道教育大学）
大橋由紀子（ヤマザキ学園大学）

本発表では，小中学校での英語授業のコーパスの構築において，インタラクションタグ付与が小中の学習指導の継続性を調べる上で有用なアノテーションとなるかの調査を試みた結果の報告がなされた。具体的には，国立大学に附属する小学校 6 年生の最後の英語の授業 4 回と連続する年度の中学校 1 年生の最初授業 6 回の授業データのインタラクション分析を行い，小中の「継続性」がどのように観察されるかについて報告がなされた。

教師・生徒の発話（日英両語）について，Walsh (2006) の言う SETT (Self-Evaluation of Teacher Talk) から 4 つの interaction mode (managerial, materials, classroom context, skills and systems)と Ellis (1984)の述べる social という考えをインタラクションタグの属性値として組み込んで分類が行われた。

分析した結果，小学校では前半は Skills and systems モードを多用し後半で classroom context で交流させるパタンが見られ，中学校では，materials モードで全体を把握させてから，Skills and systems mode, classroom context mode へと移って, classroom context で交流をさせる，最後に skills and systems mode, classroom context mode を取り入れながら，materials mode に戻るということが観察された。小中両者に共通して，skills and systems mode が重要視されている傾向があり，中学がより交流を持つことのできる授業展開（後半の classroom context の多用）となっていることが読み取れた。

このインタラクション情報により小中英語授業の一定の継続性を捉えることができたが，インタラクションモードの回数の観測だけでは判断が難しく，今後はそれぞれのモード内のターン数を考慮することが望ましいことが課題として指摘された。

第 1 日第 3 セッション
［司会・報告］後藤一章（摂南大学）

「複数コーパスの統一的処理を可能とした高速コーパスデータベースシステム MyCo の開発」

西村祐一（元名古屋大学大学院生）

コーパスからの用例検索には，主にネット上の検索サービス利用とコーパス自体を手元のサーバーないし PC に格納して行う検索の 2 方法がある。前者では検索アルゴリズムが利用者には不明であり，時にはコーパス自体が不明なこともあるなどいわばブラックボックスである。後者ではコーパスごとに独自の記録形式で作成されており，複数コーパスがある場合にはコーパス別に処理しなければならない。MyCo は，後者の場合に複数コーパスを統一的に取扱い，かつ高速検索を目標に発表者が開発したシステムである。

MyCo では，最小要素 WLP (Word+Lemma+Pos)を 1 行に編集することで，定義の異なるコーパスを統一的に扱えることに着目した。コーパス単位に WLP ファイル編集プログラムを作成すれば，あとは共通処理で同一形式のデータベースを作成できる。データーベースシステムには MySQL を採用した。なお，研究に必要な用例は検索語を NODE とする ngram であることから，NODE の前後各 10WLP からなる 21gram を編集してデータベースに格納し，高速検索の実現を企図した。BNC-XML で作成したデータベースでは，高頻度の because，

something でも 3 秒以内で用例抽出できるので利用者はストレスを感じない。抽出用例をもとにピクチャー，kwic，コロケーション表を出力できる。さらに，抽出用例はテキストファイル形式で利用者のホームディレクトリーに書き出されるので，利用者が直接処理することも可能である。

会場から，端末操作方法は Web アプリ形式ではないのかとの質問があり，コーパスを直接利用する研究者を対象にしているのでコマンドラインモードで開発したと回答した。また，Pos Tag の統一についての質問には，将来的には適切なタガーを採用して統一することを課題にしていると回答した。

「構文構造を活用した学術論文における頻出コリゲーションの抽出」

田中省作（立命館大学）
徳見道夫（九州大学名誉教授）
宮崎佳典（静岡大学）
金丸敏幸（京都大学）
田地野彰（京都大学）

本発表は，多くの学術分野で頻出する表現の抽出法の提案である。まず，コーパス研究でよく使われる n-gram でコーパスから表現を切り出すと，自由項を挟むような不連続な頻出表現が得られないこと，切り出された表現に分野共通の表現とそうでないものが混在してしまうことが問題点として指摘された。提案手法は，これらの問題を解決したもので，主に次の 2 点で画期的である。

まず，構文情報の導入である。論文コーパス中の特定分野で使用される語をエントロピーで同定し，予め痕跡化しておく。この痕跡を周縁の非内容語等も含めつつ，適切な中間情報にするが，それに元の英文の構文構造を活用している点である。その結果，構文情報が元の英文に埋め込まれる。

また，表現の切り出しには n-gram を用いるが，元の英文（コーパス）が前述の手続きで事前に適度に抽象化されていることから，不連続な頻出表現も抽出が可能となっている。元の英文における通常の n-gram との対応数による n-gram の重み付けも巧みで，一部の頻出表現は文法的な情報が埋め込まれたコリゲーションとして得られる。異なる n でも提案手法を併用することが可能であることも述べられ，今後の展開が楽しみである。

なお，発表後には特定分野の語を同定する際のエントロピーに対するしきい値の根拠や，その他の指標に関する質問があり，前者は暫定的なものであること，後者についてはそれぞれに特徴があり，唯一的な選択が難しい旨が述べられた。

「Word2Vec による文学作品の時代比較：コーパスを軸とした異分野融合研究の試み」

内田諭（九州大学）
下條恵子（九州大学）
渡邉智明（九州大学）
斎藤新悟（九州大学）
谷口説男（九州大学）

本研究の目的は，Word Embedding の代表的な手法の一つである Word2Vec を用いて独自に作成した文学コーパスにおける単語の用法を検証し，文学研究への応用の可能性を探ることである。この手法の特徴は，共起関係に代表される syntagmatic な関係にある単語ではなく，類似のベクトルを持つ単語を探索することで paradigmatic な関係にある単語を検証することができるという点である。特定の単語と類義的に使用されている単語を調査することで，その単語が使用される文脈や含意などを明らかにすることが可能となる。

本研究では 1900 年代前半の文学作品からなる「前半コーパス」（約 180 万語）と 1900 年代後半からなる「後半コーパス」（約 171 万語）を作成した。対象となる作品については，アメリカ文学作品の中から，金融業界を取り扱った作品及び現実を写実的に描いたとされるリアリズム小説を中心に選定した。20 世紀のアメリカは 1945 年の第二次大戦終戦を機に国際政治の舞台で超大国の地位を獲得しただけでなく，経済面でも規制緩和を繰り返して活性化を図るなど，社会的に大きな変化を経験している。そしてこのような変化は文学作家たちの意識形成に影響を及ぼしており，作品内容だけでなく語彙レベルでその変化が生じていると考えられる。

これらの 2 つのコーパスを入力としてそれぞれの Word2Vec のモデルを構築した。その後，単語の頻度表から高頻度の名詞を選定してコサイン類似度から同義的に使用されている単語のリストを生成した。その結果，類義語のリストは年代によって興味深い違いを示すことが明らかになった。例えば，money は「前半コーパス」では pay, work, dollars などと類似度が高い。一方，「後半コーパス」では interest, market, bonds などが類似度の高い語としてリストされた。これは「前半コーパス」の時代には労働に対する具体的対価として金銭が語られているの

に対し，「後半コーパス」では金融商品など利益を生む無形の財やサービスとして語られているということを示唆する。文学研究における観察を裏付ける結果であり，コーパスを基軸とした異分野融合の有効性が示されたといえる。

質疑応答では，単語類似度計算におけるベクトル空間の次元数に関する質問があり，先行研究に基づいて設定したと回答があった。

「構文情報などを表す木構造の配列による情報処理」

田中省作（立命館大学）
宮崎佳典（静岡大学）
田辺利文（福岡大学）
田村昌彦（立命館大学）

近年，構文解析器の高度化や構文情報付きのコーパスの整備が進んでいるにもかかわらず，構文情報はコーパス研究ではあまり活用されていない。構文情報は「木構造」とよばれるデータ構造で実装されることが一般的であるが，文系出身の研究者が多い学界では取り扱いが難しいのであろう。本発表は，語列照合程度のプログラミング知識を前提とした，より手軽で分かりやすく構文情報を取り扱うための方法が提案された。

提案手法は 2 本の「配列」で構文情報を表現する。配列はほとんどのプログラミング言語でも備えている基礎的なデータ構造で，プログラミング初学者にも馴染みやすい。発表では，「木構造」に初めて出会う出席者を前提に，木構造の基本的概念から入り，構文情報の配列表現や処理がグラフィカルに分かりやすく示された。本発表の内容が言及・理解すべき概念が少なくなく，限られた時間であったため，実際のコーパス検索事例については配布資料による説明にとどまったが，その有効性が十分に感受された。発表終了後にも出席者から分かりやすい説明という感想や，「こういった文法情報は取れるか」といった質問が多く聞かれ，チュートリアルといった形で再度披露されることが期待される。

発表時の質疑では，準備されているライブラリに関すること（括弧表現から配列表現への変換）や，配列による木構造の表現に対する健全性に関する質問があり，前者については既に API として実装されていること（配布資料にも記載），後者については健全である旨の回答があった。

第１日第４セッション
［司会・報告］石川有香（名古屋工業大学）

「強調語の調査による Popular Music の歌詞の文体研究」

渡部文乃（京都大学大学院）

本発表では Popular Music の歌詞の特徴を文体の観点から報告した。本研究は，発表者が構築したアメリカ人歌手の英語歌詞コーパス American Popular Music Corpus of English (PMCE-US)のデータと，広範なジャンルから構成される Manually Annotated Sub-Corpus (MASC)を用い，強調語のもつ機能や性質を利用して，歌詞の内容，対人関係，そして媒体の特徴について報告した。その際 Ito and Tagliamonte (2003)に倣って，形容詞の意味を高める表現に限定した。

まず，ジャンルの主観性や話し手の聞き手への関わり度を調査するために，強調語の全体頻度を観察した。歌詞には 1 万語あたり 48.8 個の強調語が生じており，この頻度は MASC のどのジャンルよりも高かった。歌詞の強調語の半数が「さび」の部分に生じていたことが，その原因と考えられる。

また，ジャンルの表現の明確さやフォーマリティーを調査するために，各強調語の頻度を観察した。歌詞の強調語は Type Token Ratio の値が小さく，強調語の約 8 割が強調語 so であった点で，MASC のどのジャンルよりも際立っていた。発表では，音楽の拍が短い語の短い音節を好む傾向があることが so の多用の原因ではないかと考察した。

質疑応答では，歌詞の繰り返し部分の処理方法や「主観性」の捉え方に関する質問があった。それに対して，繰り返しも含め，すべての歌詞を数える方式を採用したことや，作詞者・歌手・聴衆という関係の中では，作詞者が歌手にことばを話させるという独特の発話形式を持っていることから，歌手を主体とみなしたとする回答があった。今後の研究課題として，強調語の対象を広げて調査を行ってみてはどうかとするコメントが出るなど，大いに議論が盛り上がった。

「ホテルのオフィシャルウェブサイトにおける概説文のストラテジー：Move の構築と分析を中心に」

近藤雪絵（立命館大学）

本研究はロンドンのホテルのオフィシャル

ウェブサイトに掲載された概説文を Swales（1990）が提唱したジャンル分析の手法を用いて分析し，読み手にアピールするストラテジーを探求することを目的として行われた。

コーパスとしてロンドンの 3-5 つ星のホテルのウェブサイトに掲載された概説文（計 124）が集積され，書き手の意図と特徴的な言語表現を元に，次の 3 つの Move と 3 つの Step が構築された。

Move 1: Defining self,
Move 2: Establishing features,
Move 3: Establishing connections,
Move 2-Step 1: Describing the history/ architecture,
Move 2-Step 2: Describing the location,
Move 2-Step 3: Describing the facilities

Move 1，2 は採択率が全てのグレードにおいて 8 割を超えており，自身を定義付けてから特徴を確立することがホテルの概説文の典型パターンであることがわかった。ホテルのグレードが下がると Move 2 の採択率は高まり，3-star では全ての概説文に Move 2 が採択された。一方で，Move 3 は 3-star ホテルでは Move 1 と同じ 8 割強の採択率であるが，5-star ホテルでは 5 割強にとどまった。Move 3 では二人称代名詞を用いて読み手に呼びかけたり，予約を促したりする表現が見られ，中グレードホテルではこのように読み手と関係を築くストラテジーが使われていた。今回の分析により，概説文のグレード毎の典型パターンが発見され，読み手にアピールするストラテジー違いは，高グレードホテルのテキストが持つ “luxury”感（抽象性，排他性）にあることが示唆として述べられた。

質疑応答では，高級ホテルの概説文は，広告会社が作っているので，書き手の違いが言語特徴に表れている可能性も考えられないか，という質問があった。それに対しては，書き手の熟達度の違いも一因として考えられるという回答があった。ついで，個々の事例の紹介があり，ホテルのランクによって，場所を表す形容詞や倒置の使用に異なる傾向があることが具体的に示された。また，それらの文体的効果について，活発な意見が交わされた。

「一般教書演説から見る米国大統領の関心事の変遷―トピックモデルと時代背景― 」

木山直毅（北九州市立大学）

本研究では米国の一般教書演説に対しトピックモデルの手法を用い，歴代米国大統領の関心事がどのように変化したかを解析した。その結果から大統領の関心事は主に 4 つのトピックの変遷を辿ったことが明らかになった。

まず 1790 年から 1910 年ごろまでは *I recommend* と *law*，*congress* と *necessary*，また *amendment to/of the constitution* といったコロケーションの多さから「内政」への興味であることが論じられた。この背景には，国として法や憲法の整備が最優先事項であるとし，議会に対しそれを促そうとしていたためではないかと考察した。

1890 年代後半から 1920 年代終わりまでの間，*farmer*，*industries*，*agriculture*，*production* といった語彙がワードクラウドに現れ，これらは「国内の産業」に関わりがあることを論じた。この背景として第一次世界大戦で米国が一人勝ちしたことによる特需があったことに起因することを論じた。

1930 年代から 1970 年代後半までは *war*，*force*，*freedom* といった語彙がワードラクウドで確認され，「軍事関係」のトピックであることを論じた。この背景は，第二次世界大戦や冷戦といった戦争を経験していることにあると考えられる。

1960 年代以降のトピックは *job*，*children*，*family*，*school*，*education* といった語彙が多く，これを「社会福祉」であると論じた。この背景には米国の相対的立場の低下にあることを論じた。

質疑応答では，1) 一般教書演説が文書で出されていた時期の「書きことば」の影響，2) ラジオやテレビの普及によって，目の前の議員から広く米国民へと変化した，聴衆の変化に伴う影響，3) 大統領と連邦議会の間の政治信条の違いによる「距離感」の影響をどのように捉えるのかという質問があり，データの解釈について，活発な意見交換が行われた。今回の発表では，モデルによって抽出された 35 のトピックのうち，時代の特徴が明確に示された 4 つのトピックに焦点をあてたが，同じ手法でさらに細かな分析も可能であるため，今後継続して分析を行う予定であるという回答があり，さらなる研究の発展が期待できるものであった。

第１日第５セッション
［司会・報告］ 大谷直輝（東京外国語大学）

【賛助会員発表】コーパスの示す科学的データと学習性・商品性との両立―『ウィズダム英和辞典』の編集にあたって―

井上永幸（広島大学）
西垣浩二（㈱三省堂辞書出版部）

辞書編集におけるコーパス活用の一般的利点にふれたあと，辞書編集がかかえる問題点を，時間・スペース・明解さの観点や，ターゲットユーザー・規範性と記述性・実用性・コーパス言語学・理論言語学との相性といった観点から概観した。次に，コーパスを全面的に活用した日本初の英和辞典として，初版以降 2 回の改訂を行い現在第 3 版が刊行されている『ウィズダム英和辞典』について，コーパス分析の成果がいかに紙面に盛り込まれているのか，コーパスから得たデータやそこから漏れるものを，いかにして掘り下げ学習者にとって有益な記述として作り上げてゆくのかという点について，実例をもとに論じた。具体的には，見出し語の頻度ランク表示，語義配列，用例配列，注記・コラムにおけるコーパス分析情報，文法的・語彙的コロケーション情報などが，どのように紙面に反映されているかを示し，編集段階からの電子データによる編集が刊行とほぼ同時に Web 版辞書「DUAL ウィズダム」を公開することに貢献していることを紹介した。また，学習辞書編集においては，コーパスの産出する科学的データはそれ自体が目的となるのではなく，それをもとに学習者にとって必要な情報を取り出し，いかに学習効率が上がるように情報を配置してゆくかという点が，学習者にとって使いやすく，結果的に商業的にも成功する教材を編纂するにあたって肝要な点であることを提示した。

発表の後の質疑応答では，フォーマル度の高い語の場合，コーパスの用例が難しくなってしまうがどのように対応するかという問いや，紙の辞書と電子辞書において，コーパスの役割が違ってくるか等の質問が出て，辞書作りにおけるコーパスの役割が議論された。辞書編纂者の辞書にかける思いが伝わってくる発表であった。

「英語辞書レーベルとコーパス」

田畑圭介（神戸親和女子大学）

本発表では英語学習辞典が採用するレーベルの体系化と細緻化の必要性を検討した。レーベルはその特性上，ヘッジが付与されうる集合タイプと付与すべきでない段階的タイプに二分できる。*Oxford Advanced Learner's Dictionary* (*OALD*)の foreigner (*sometimes offensive*)のレーベルは該当語の使用に注意を要することを学習者に喚起でき，[disapproving]が付される語との心的態度の相違を示すことができる。dinky のように英米で意味が正反対となる語も，*OALD* のように(*British English*, *approving*) small and neat in an attractive way，(*North American English*, *disapproving*) too small，と提示することで学習者に各語義の相違を明解に伝達できる。レーベルの体系化と細緻化は英語教育の分野においても重要なトピックとなる。また *Oxford Learner's Thesaurus* で[rather formal]と記される fortunately と[especially spoken]と記される luckily の両語の頻度差から，話し言葉コーパスの formality が計れる可能性についても言及した。luckily よりも[rather formal]の fortunately を多く含む話し言葉コーパスを順に配列すると，COCA Spoken Section＞TED Corpus＞テレビドラマ HOUSE＞テレビドラマ Desperate Housewives＞COLT となる。

発表後の質疑応答では，様々な抽象性の段階で記述が可能なレーベルにおいて，どの程度の抽象性のレーベルが妥当であるかを決める客観的な基準があるのかという質問がなされた。個々の例において最も適切なレーベルを決定するのは可能であっても，レーベルの体系的な設定は大変な作業であることが実感できる発表であった。

「怒りを表す類義語と概念メタファー」

南澤 佑樹（大阪大学大学院生）

怒りの感情に関しては，概念メタファー理論の枠組みに基づき数多くの研究が行われてきた。先行研究では，怒りに用いられる主要なメタファーが複数提案され，その中でも ANGER IS A HOT FLUID IN A CONTAINER は最も中心的なものとされている(Kövecses 1990, 2000)。しかし，類義語における概念メタファーの結びつきの違いに関しては，従来あまり関心が向けられていなかった。

コーパスを用いたメタファー分析に関しては Stefanowitsch (2006)が根源領域に属する語と目標領域に属する語が共起する表現を分析対象とする手法を提案しており，この手法を用いて鈴木（2010）や Turkkila (2014)が類義語に見られる概念メタファーについて分析を行っている。しかしコーパスを用いた分析では，メタファー

表現の出現頻度を重視するために boiling with anger のような感情の様態を具体的に表す表現よりも in anger のような意味内容の薄い表現を中心的とみなしてしまうといった問題も見られる。

したがって本発表では，British National Corpus より MI スコアを用いて anger, rage と結びつきの強いメタファー表現を収集した。それらを概念メタファーごとに分類した結果，anger は vent, seethe, well などと結びつきが強く ANGER IS A HOT FLUID IN A CONTAINER が最も中心的な概念メタファーであった。一方 rage では howl や bristle といった語が上位にあがっており ANGER IS A DANGEROUS ANIMAL とも結びつきが強い。また anger と結びつくメタファー表現は怒りの様々な側面を表すのに対し，rage と結びつきの強いメタファー表現は怒りの暴力的な側面を表す傾向が見られた。さらにメトニミー的に怒りを表す共起語の観点からも以上の結果が支持された。

発表後の質疑応答では，BNC の調査に対して，話し言葉と書き言葉を分けて調査をすることで，見えてくるものもあるのではといった質問や，怒りのメタファーの中でも中心的なものと周辺的なものがあるのではという意見が出た。コーパスの詳細な記述の理論的な解釈をしっかりすることで，実証的なメタファー研究になるように思われた。

第１日第６セッション
［司会・報告］森下裕三（環太平洋大学）

「日英対訳コーパス中の「〜ことになる」構文とその英訳文間の構造的不一致」*

大矢 政徳（目白大学）

機械翻訳における BWA (Bitext Word Alignment)と呼ばれる手法には，いくつかの問題点があると言われている。この手法は，翻訳元の語を翻訳先の語に対応させるという発想が基盤にある。しかし，実際には対訳コーパスなどでも翻訳元の語が翻訳先の語と対応関係にないことも少なくない。たとえば，日本語の「〜ことになる」は英語で“be supposed to”と対応していると考えられる。だが，実際には日本語の「〜ことになる」が常に“be supposed to”と翻訳されるわけではなく，語と語の対応関係を探るだけではうまくいかない。

このような BWA という手法の不備を補うべく，統語的な依存関係の不一致をもちいた新しい手法が機械翻訳において有効であることを示す。まず，日本語と英語における統語的な依存関係がどのくらい一致しないのかを示すために，対訳コーパスとして知られている『Wikipedia 日英京都関連文書対訳コーパス』を使用する。このコーパスには，日本語の「〜ことになる」を含む日本語の用例が 1,197 例含まれており，これらのうち，ランダムにサンプリングした 100 例を調査対象とする。この 100 例の日英語対訳データの統語的な依存関係を詳細に確認したところ，特に「〜ことになる」の部分が依存木の根の位置に見られる場合に，統語的な依存関係が日本語と英語で一致しないことが多いという点を明らかにした。どのような構文で，日英語の統語的な依存関係が異なりやすいのかを明らかにすることは，機械翻訳の研究において大きな意義がある。

質疑では，機械翻訳という分野そのものにかかわる技術的な問題点などが議論された。既に機械翻訳では技術的に困難な問題が多く指摘されており，今後，さらに研究の発展が期待されている分野である。本研究はそうした機械翻訳という分野における研究の可能性を示す研究のひとつであった。

（要旨も司会者が執筆）

「医学研究論文ジャンルにおけるコーパス作成ツール AntCorGen を活用した教育の可能性−Construction of Corpora for Discipline-Specific Learning in Medical Research Article Genres」

浅野元子（大阪大学大学院生）

AntCorGen (Anthony, 2017)と呼ばれるコーパス作成ツールをもちいた PLOS ONE 誌の医学研究論文コーパスを構築する試みと，データ駆動型学習 (DDL; Lee and Swales, 2006)への応用に際して専門領域の論文を限局する試みについて報告した。

まず，AntCorGen によって構築された Cardiology（心臓病学），Gastroenterology and hepatology（胃腸病学と肝臓学），Pulmonology（呼吸器学），Oncology（腫瘍学）の各分野から約 5,000 報の論文からなるコーパスを準備した。さらに，コーパスから各分野の論文を 100 報ずつ無作為に抽出し，この無作為抽出された 400 報の論文からタイトルや抄録でのムーブ（Swales, 1990; Salager-Meyer, 1990 and 1992）とヒント表現（Tojo, Hayashi and Noguchi, 2014）によって症例コホート研究などの研究の種類

（国立国際医療センター, 2009）に分類した。本文の総語数は約 173 万語で，異なり語数を総語数の平方根によって割ることで算出できる Guiraud Index は 36.8 であった。また，高頻度後を変数としてクラスター分析（田畑, 2014）を行うと，動物とヒトでの研究に大別され，研究の種類による類似性が示唆された。この結果から，本ツールは学術論文コーパス構築に有用であることが判明した。本ツールによって構築されたコーパスから医学生や研究者が行う研究と同一種類の研究論文を選択して各自のコーパスとすることで，目標とする専門家集団が慣れ親しんだ修辞パターンを教育現場でもより実戦的に学習することができると考えられる。

質疑では，AntCorGen によるコーパス構築について複数の質問が寄せられ，AntCorGen という新しいコーパス作成ツールへの関心の高さが伺われた。また，ジャンル分析の一種であるムーブ分析についても活発な意見の交換がなされ，学術論文におけるムーブ分析への期待の大きさも感じられた。

「Applying Topic Models to Describe the Composition of the FLOB Corpus: How can the external criteria be associated with meaningful sets of internal evidence?」*

Tomoji Tabata (Osaka University)

トピックモデルとは，テクスト中に隠された意味の構造を見つけ出す機械学習の手法である。伝統的な統計的手法とは異なり，トピックモデルでは複数のテクストに含まれるキーワードを対象とした分析が可能である。さらに，関連性の強いテクスト同士を分類することができ，この結果をヒートマップなどのさまざまな形で可視化できるというメリットもある。

本研究では，FLOB コーパスを使用して，これまでの他の研究手法ではうまく扱えないと考えられてきた低頻度語に注目した分析を行った。FLOB コーパスのトピック構造を分析した結果，情報散文と小説などの架空の事柄について書かれた文章はトピックモデルで明確に区別されることが明らかになった。さらに，レジスター間でのトピックの分布を調査してみたところ，トピックそのものも情報散文と関連性の強いものと小説のような架空の事柄について書かれた文章と関係が強いものに分類可能であることが明らかになった。本研究の分析結果は，テクスト中における意味のパターンを解き明かす新しい試みとなることが期待される。同時に，機械学習によるテキストマイニングから得られた知見と伝統的な文体研究との架け橋になる可能性も示唆される。さらに，トピックモデルが，機械学習をもちいた最先端の技術であるだけでなく，扱うのが困難だと考えられてきた意味の分野を対象にした研究でもあることを示すことができた。

質疑では，トピックモデルという新しい手法について活発に議論が行われた。まだ，コーパス言語学の分野でも十分に知られていない最新の手法であるが，トピックモデルには多くの関心が寄せられていることがよく分かる議論となった。トピックモデルによる研究が今後もさらに増えていくことが望まれる。

（要旨も司会者が執筆）

講演，シンポジウム，ワークショップの要旨は各登壇者が執筆しました。

■講演

「A Frontier in Learner Corpus Studies: For Better Understanding of L2 Learners」

Shin’ichiro Ishikawa (Kobe University)

Various learner corpora have been developed to date and they have greatly contributed to improvement of L2 teaching. However, a more carefully designed corpus would be needed for a reliable contrastive interlanguage analysis. Thus, recent learner corpora have come to pay much more attention to controlling variety in the collected data.

The International Corpus Network of Asian Learners of English (ICNALE) is one of the largest learner corpora ever compiled. It includes more than 10,000 speeches and essays produced by L2 English learners in ten countries and regions in Asia as well as English native speakers. Its unique feature is that the topics are carefully controlled. All the participants are required to speak or write about two kinds of common topics: (A) It is important for college students to have a part-time job and (B) Smoking should be completely banned at all the restaurants in the country. Such a topic control is expected to lead to a greater reliability in 19 varied types of contrastive analyses (Ishikawa, 2013).

The ICNALE currently consists of four modules: Spoken Monologue (1,100 participants, 4,400 samples, 500,000 tokens), Spoken Dialogue (under construction), Written Essays (2,800 participants, 5,600 samples, 1,300,000 tokens), and Edited Essays

(290 participants, 580 samples, 140,000 tokens).

The ICNALE development team believes that comparing something comparable is a key to further development of learner corpus studies.

■シンポジウム

「話し言葉コーパスの構築と利用」

司会：野口ジュディー
（神戸学院大学名誉教授）

本シンポジウムでは，各講師が構築したコーパス〔学習者の書き言葉・話し言葉（英語）〕ICNALE，学習者話し言葉日本語（テーマ別）I-JAS，話し言葉日英（理系プレゼン）JECPRESE，TED コーパス〕に関して，構築とその利用方法についての話がなされました。以下，各講師によって書かれた要旨を掲載します。

「The ICNALE：中間言語対照分析の精緻化とアジアにおける学習者コーパス研究の発展を目指して」

石川慎一郎（神戸大学）

The ICNALE (The International Corpus Network of Asian Learners of English)は，アジア圏 10 か国・地域において，英語学習者の L2 産出データを収集するプロジェクトで，すでに，Written Essays，Spoken Monologue，Edited Essays の 3 つのモジュール（計約 190 万語）が公開され，現在は，Edited Essays モジュールの拡充と，初のマルチモーダル版となる Spoken Dialogue モジュールの開発が進められています。The ICNALE の特徴は，プロンプトやテキストの長さなどが一定の範囲で統制されていることで，これにより，信頼性の高い国際比較研究が可能となります。The ICNALE は，ダウンロード版のほか，オンライン版があり，専用の検索用インタフェースが開発されています。

「International corpus of Japanese as a second language：日本語学習者の言語研究と指導のために」

迫田久美子（広島大学・国立国語研究所）

International corpus of Japanese as a second language (I-JAS, http://lsaj.ninjal.ac.jp/)は，12 の言語を母語とする日本語学習者の発話と作文のコーパスです。JFL 学習者と JSL 学習者，さらに国内の学習者は教室環境と自然環境学習者のデータが収められ，さらに同じタスクを実施した日本語母語話者のデータも含まれています。完成は 2020 年春を予定しており，現在は 450 名のデータが公開されています。I-JAS の特徴としては，英語，中国語，韓国語，西語，独語，仏語，露語，タイ語，トルコ語，インドネシア語，ベトナム語，ハンガリー語の母語話者の複数のタスクのデータを所収しており，検索システムを備えていることが挙げられます。全員が同じテストを受けており，成績や背景事情も公開しています。

「JECPRESE：JSL と EFL ユーザーのために」

野口ジュディー（神戸学院大学名誉教授）

JECPRESE, the Japanese-English Corpus of Presentations in Science and Engineering (http://www.jecprese.sci.waseda.ac.jp/)は留学生のための専門日本語教育（JSL, Japanese as a second language）を支援する研究発表コーパスでスタートしました。日本の大学院生の日本語プレゼンテーションに加えて，アメリカの大学生や国際学会の英語プレゼンテーションも収められていて，EFL (English as a Foreign Language)の学生にも利用できるコーパスになりました。理工系のプレゼンテーションの特徴をわかりやすくするために，各発表を ESP (English for Specific Purposes)の手法であるジャンル分析に基づいてセクションやステップで検索できるようにしました。単語や表現の検索もできます。

「TED Corpus Search Engine：TED Talks を研究と教育に活用するためのプラットフォーム」

長谷部陽一郎（同志社大学）

本発表では TED Talks の英語トランスクリプトを検索するための Web システム TED Corpus Search Engine (TCSE)を紹介し，本システムが多様な言語研究・言語教育プロジェクトで活用できる可能性があることを示した。TCSE に収録されている言語データには次のような特徴がある。1) テキストと同期された音声・動画データを持つ。2) 話し言葉ならではの表現を多数含む。3) 概ね 18 分以内の，ある程度統制された形式のトークで構成されている。4) 多言語による対訳データが利用可能である。こうした特徴は，英語の分析や教育に関わる研究者が，様々な口語表現や談話標識などの意味や機能を分析するのにとりわけ役立つであろう。また本発表では，TCSE が提供するデータ及び機能の認知言語学な価値についての考察を示した。用法基盤の立場を取る認知言語学では，個々の発話場面における一回的な経験が抽象化されるこ

とで，語彙や文法の構造が起ち上がってくると考える。TCSE を用いて採取可能な言語データは TED Talks という形式の範囲内で完全な文脈の再現が可能であり，その意味において，発話場面情報の再現可能性を必ずしも重視しない従来型のコーパスと趣を異にしていると言えよう。

■ワークショップ 1

「TCSE を用いた TED Talks の全文検索と英語教育への応用」

長谷部陽一郎（同志社大学）

TED Corpus Search Engine (TCSE)は TED が公開している約 2,400 件の英語プレゼンテーションのトランスクリプトを解析してデータベースに格納し，英語テキストと翻訳テキストの全文検索を可能にした Web システムである。TCSE には英語教育や言語学研究のために TED Talks を役立てるための各種機能が実装されており，本ワークショップでは特に英語教育での活用を念頭においた解説を行った。その構成は概ね次の通りである。まず TCSE に収録されている TED Talks データの概要や基本統計情報を示した。次に TCSE の基本的な使い方を説明した後，TCSE に実装されている検索シンタックスを使った高度な検索（レンマ検索，品詞検索，ワイルドカード検索，否定一致検索など）について解説した。また，独自の方法で TED Talk の動画を表示し，学習者のリスニング／スピーキング力を伸ばすための「ポーズ・アンド・チェック機能」を紹介した。最後の質疑応答では，今後のデータの拡充や機能の追加などについて，参加者から要望や質問が寄せられた。なお，本ワークショップの配付資料 PDF は次の URL でダウンロード可能である。

https://goo.gl/DC24tf

■ワークショップ 2

「機械学習を用いたコーパス分析入門」

小林 雄一郎（日本大学）

本ワークショップでは，近年コーパス言語学の分野でも盛んに利用されるようになってきた機械学習（machine learning）の技術を紹介しました。機械学習は，人間が持つ学習能力をコンピュータに持たせることを目指す人工知能の研究分野です。また，コンピュータにデータを解析させることで，データの背後に潜むパターンを発見（学習）させる技術のことを指します。そして，多くの場合，データから発見されたパターンは，新たなデータの予測に活用されます。機械学習の技術を用いることで，手作業では扱えないような大量のテキストデータを効率的に分析できるようになります。そして，パターンを発見するための十分な量のデータを用意すれば，人間が予測するよりも高い精度で予測を行うことが可能になります。さらに，予測に寄与したパターンを吟味することで，分析対象のテキストを特徴づける言語項目を特定することができます。

コーパス言語学における機械学習の活用事例としては，テキストの著者推定やジャンル推定，英作文の自動採点，語彙や文法の使用に関する通時的分析などがあります。本ワークショップでは，このような事例を紹介しつつ，機械学習の基本を講義形式で詳しく説明しました。ワークショップの流れとしては，(1) 機械学習とは何か，(2) データの準備方法，(3) 具体的な仕組みと手順，(4) 分析結果の検証方法，(5) コーパス言語学における活用事例，というものでした。

なお，質疑応答では，「様々な機械学習の手法の中から，どのように最適な手法を選ぶか？」，「不均衡なデータをうまく分析するにはどうしたらよいか？」，「具体的にどのような形式で頻度を集計すればよいか？」などの質問が寄せられました。

＜第 44 回大会発表者募集＞

英語コーパス学会第 44 回大会は，2018 年 10 月 6 日（土）と 10 月 7 日（日）に東京理科大学で開催されます。（神楽坂キャンパスで開催できるよう調整していますが，変更になる可能性があります。）例年通り研究発表を募集いたしますので，発表を希望される方は，下記の要領に従い奮ってご応募下さい。

【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。未発表の研究に限る。
【応募資格】本学会員であること。（連名発表の第二（以降）発表者は必ずしも会員でなくても構わない。）同一人物が代表者となる複数件の発表申し込みは認めない。（自身が代表者である発表申し込みとは別の発表申し込みで連名発表の第二（以降）発表者となることは妨げない。）
【発表方法】研究発表（発表 20 分，質疑 10 分）
【応募方法】発表申込ウェブフォーム（https://goo.gl/forms/rYhmgqUiQ6brgezs1）に必

要事項を記入の上，発表概要を事務局長（jaecs.hq@gmail.com）宛に電子メール添付ファイルで送付してください。

学会ウェブサイトの大会等のページにあるテンプレート（Word 形式）をご利用の上，Word 形式のままで送付してください。

発表概要は冒頭に題目のみを記し，概要本体（題目・文献リストを除く）を 800～1,200 字でお書きください。必要に応じて参考文献を明示してください。ただし，文献リストの部分は前述の文字数の集計対象外とします。

ツール・コーパス開発などの発表を除き，リサーチクエスチョンならびに研究から得られた知見を明快に記述してください。

発表概要は応募者が容易に特定されることのないようご留意の上作成してください。同様に，応募者推定につながる文献は挙げないでください。

脚注は使わないで下さい。

発表は英語でも構いません。その場合は，概要は 400～600 語でお書きください。題目・文献・氏名・所属等も全て英語でご記入ください。それ以外は上記と同じです。

メール本文には代表者名と発表題目を明記してください。

※応募情報は審査終了まで事務局長のみが扱い，審査は発表概要のみに基づいて行われます。

※発表が採択された場合には，大会資料に掲載する要旨（文献リストは掲載しません）と，大会後に発行されるニューズレターに掲載する報告（大会資料の要旨を実際の発表に基づいて適宜修正したもの）を執筆していただきます。

※発表概要から容易に応募者が特定されると考えられる場合には，応募書類を加工して審査に付する可能性があることを予めご了承ください。

【応募期限】2018 年 5 月 31 日（木）必着
（※今回から期限が 1 か月早まりましたのでご注意下さい。）
【採否決定】2018 年 7 月初旬（予定）

＜理事会の決定事項について＞

9 月 29 日（金）17 時 30 分より関西学院大学において理事会が開催されました。承認された人事についてご報告いたします。

(1) 会長・副会長・事務局長・事務局補佐

任期が 2017 年度末までの会長・副会長・事務局長・事務局補佐の任期について，もう一期の継続が承認された。

(2) 理事

・理事（新任）

アントニ ローレンス先生（早稲田大学）
金澤俊吾先生（高知県立大学）
小島ますみ先生（岐阜市立女子短期大学）

(3) 編集委員会

・編集委員（新任）

田畑智司先生（大阪大学）

(4) 大会企画委員

・企画委員（新任）

小島ますみ先生（岐阜市立女子短期大学）

＜会則改訂について＞

総会において下記の 2 回分の会則改定が承認されました。

(1) 第 16 条を改定し，東支部を廃止し，研究会を設置する。研究会を今年度より運用するための遡及適用を可能とするために改定日を 2017 年 4 月 1 日とする。

(2) 第 5 条の細則にある理事の年会費を廃止し，平成 26 年度から一時的に 10,000 円に値上げしていた理事の年会費を一般会員と同じ 6,000 円に戻す。改定日は 2018 年 4 月 1 日とする。

＜研究会（SIG）の発足について＞

総会において会則改訂が承認されたことにより，先の理事会で承認された下記の 5 つの研究会が正式に発足しました。

ESP 研究会（JAECS SIG on ESP）
代表：石川有香先生（名古屋工業大学）
副代表：藤枝美穂先生（大阪医科大学）

DDL 研究会（JAECS SIG on DDL）
代表：中條清美先生（日本大学）
副代表：佐竹由帆先生（駿河台大学）

コーパスと CEFR 研究会（JAECS SIG on Corpora and CEFR）
代表：投野由紀夫先生（東京外国語大学）
副代表：内田諭先生（九州大学）

語彙研究会（JAECS SIG on Lexicology）
代表：石川慎一郎先生（神戸大学）
副代表：杉森直樹先生（立命館大学）

ツールと統計手法研究会（JAECS SIG on Corpus Tools and Statistical Methods）
代表：アントニ ローレンス先生（早稲田大学）
副代表：水本篤先生（関西大学）

各研究会への入会，研究会の新規設立は随時

受け付けています。詳しくはウェブサイトの研究会のページをご覧ください。

＜会誌『英語コーパス研究』第 26 号論文投稿募集について＞

『英語コーパス研究』編集委員会委員長
中尾佳行（福山大学）

『英語コーパス研究』第 26 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」,「総説論文」,「書評論文」,「実践報告」
2. 「書評」,「コーパス紹介」,「ソフトウェア紹介」,「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
【原稿提出期限】2018 年 11 月 30 日（金）
電子メール添付にて提出してください。提出方法等についての詳細は学会 Web ページの投稿規定 http://jaecs.com/jnl/jnl_kitei.pdf を参照してください。
【問い合わせ先・原稿提出先】
『英語コーパス研究』編集委員会
E-mail：jaecs.ed@gmail.com
【採用通知】2019 年 1 月
【発行日】2019 年 3 月 31 日
（発送は 2019 年 5 月下旬の予定）

＜英語コーパス学会 学会賞・奨励賞の募集について＞

英語コーパス学会 学会賞選考委員会 委員長
西村秀夫（三重大学）

2018 年度英語コーパス学会賞および奨励賞を募集いたします。学会賞は，英語のコーパス利用を中心に据えた英語研究・教育，あるいはその関連領域の研究や学会活動などに，多大な貢献が認められる業績に対して贈られる賞です。今までに，著書，一連の複数論文，コーパス分析ツールの開発などの業績に対して授与されています。同時に，特に若手研究者を対象に，奨励賞も募集します。こちらは，若手研究者の優れた業績に 報いるために設けられた賞です。どちらの賞の応募期限も，2018 年 6 月末日です。奮ってご応募ください。

【対象】学会賞は，英語コーパス学会の目的に照らし，英語コーパスに関わる特に優れた研究業績（著書，一連の複数論文，コーパス分析ツールの開発，その他）をあげた学会員（個人またはグループ）とする。奨励賞は，39 歳以下で，英語コーパスに関わる優れた研究業績（著書，学会誌『英語コーパス研究』に掲載された論文 1 編以上，コーパス分析ツールの開発，その他）をあげた学会員個人を対象とする。
【応募方法】自薦，他薦を問わない。
【提出書類】1）推薦理由書（所定の書式（Word または PDF）による。学会ウェブサイトの学会賞のページからダウンロード可能。）単行本の場合：事務局で用意するので送付は不要。論文の場合：現物またはコピーを送付。
※ネットから自由にダウンロードできるものは，ダウンロード先の明示のみでよい。
※奨励賞対象が論文の場合は，『英語コーパス研究』に限定されるので送付は不要。

【提出先】英語コーパス学会 学会賞選考委員会 委員長 西村秀夫
E-mail：jaecs.award@gmail.com
【応募期限】2018 年 6 月 30 日（土）
【審査結果の報告および表彰式】第 44 回大会総会（東京理科大学 10 月 6 日）

＜2018 年度春季シンポジウムについて＞

2018 年 4 月 21 日に東京外国語大学にて春季シンポジウムを開催する予定です。詳細については決まり次第，学会ウェブサイト，メーリングリストにてお知らせいたします。

＜今後の大会日程と開催校＞

第 45 回大会は 2019 年 9 月下旬から 10 月上旬に高知県立大学にて開催する方向で現在調整を行っています。

＜新入会員紹介＞

寺島芙由	（東京外国語大学, S）
Du Xiangtao	（東京外国語大学, S）
小藤晃裕	（東京外国語大学, S）
田金雄一	（国際教養大学）
南澤佑樹	（大阪大学, S）
浦和千恵	（香港理工大学）
井本美子	（放送大学）
武藤明弘	（愛知大学）
山崎加奈	（東京外国語大学, S）
中西　淳	（神戸大学, S）
山本史歩子	（青山学院大学）
福嶋祐貴	（京都大学, S）

張　晶鑫　　（神戸大学, S）

（S は学生会員）

（2017 年 7 月 2 日から 2017 年 12 月 1 日の入会者）

＜事務局から＞

事務局からは情報発信のツールとして，学会ウェブサイト，ニューズレター，メーリングリストなどでイベントの案内などを随時行っております。

◇会費納入のお願い

2017 年度会費（一般 6,000 円，学生 3,000 円）未納の方は，6 月または 9 月にお送りした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：00930–3–195373〕。払込取扱票を紛失された方は，郵便局に備え付けのものに加入者名「英語コーパス学会」とご記入の上お納めください。

過年度会費未納の方は，2017 年度分と併せてお納めください。過年度会費未納の場合，機関誌などの送付を一時中止させていただいております。

住所，所属などに変更や異動のある方は，学会ウェブサイトの「会員情報変更」からのお手続きをお願い申し上げます。

※会員の皆様には，日頃より会費の当該年度内納入に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，当該年度内に学会ウェブサイトの「退会手続」からのお手続きをお願い申し上げます。

☆☆☆☆☆☆☆☆☆☆

FORUM

■4th Learner Corpus Research Conference (LCR 2017)に参加して

三浦愛香（東京農業大学）

ベルギーの Sylviane Granger 氏らが中心となって立ち上げた学習者コーパスの国際大会 Learner Corpus Research も本年の開催で第 4 回目を迎えた。2017 年は，イタリアの Bolzano/Bozen にて 10 月 5 日から 7 日に開催された。78 名の発表者と 126 名の参加者があったという。

基調講演は，異なる観点から学習者コーパスを捉えた内容であった。第一日目にペルージャ外国人大学の Stefania Spina 氏による中国人イタリア語学習者による縦断的コーパス Longitudinal Corpus of Chinese Learners of Italian (LoCCLI)についての講演，第二日目は，エクセター大学の Philip Durrant 氏による英国における 6 歳児から 16 歳までの第一言語のライティングをコーパス化した縦断的な研究，第三日目は，カリフォルニア大学サンタバーバラ校の Stefan Th. Gries 氏による学習者コーパス研究における量的手法の重要性についての講演であった。Spina 氏は，SLA と言語教育の融合や質的分析のデザインの重要性を，また Durrant 氏は，手書きの作文の読み取りの難点などコーパス構築における課題等について述べていた。また，Gries 氏は，いくつかの先行研究結果を取り上げ，回帰モデルを用いてどう改善できるかというおなじみのケーススタディーを示してくれた。はずれ値や値の散らばり，様々な変数の影響は，学習者コーパスの研究で見逃しがちだが，時には誤った結果を導く可能性があると丁寧に説明してくれた。ただし，あまりにもテクニカルな発表に，「学習者コーパスを研究し活用する利用者の多くは，言語研究者というよりは教育者であるから，よりわかりやすく汎用性のある統計手法を提案して欲しい」という聴衆からのやや批判的なコメント及び質問の応酬は，非常に興味深いものとして拝聴した。

学会では，Full paper と Work in Progress の研究発表の他，ポスター発表やソフトウェアのデモンストレーションが行われた。研究発表のテーマは，縦断的研究，特徴分析，自然言語処理，教授法，習得・教育，アノテーション，翻訳，複雑さ，10 代の学習者，アカデミックコーパス，評価，話し言葉，コロケーションなど多岐に渡っていた。

英語コーパス学会の会員による発表は，以下である（発表順に記載）。特徴分析のテーマセッションにて，“Extraction of unsuitable pragmalinguistic features of requests produced by Japanese learners of English with low proficiency” (Aika Miura)，自然言語処理のテーマセッションにて，“What kind of linguistic features distinguish second language learners’ texts from those of native speakers, and why?” (Masatoshi Sugiura, Daisuke Abe and Yoshito Nishimura)，そして習得・教育（WIP）にて，“Tracking L2 language development through construction of a longitudinal spoken learner corpus” (Mariko Abe, Yasuhiro Fujiwara and Yuichiro Kobayashi)の発表

があった。名古屋大学の杉浦正利先生そして中央大学の阿部真理子先生による発表は，自然言語処理と縦断的学習者コーパスの構築という最先端の研究に関するもので，会場の席が足りなくなるほど多くの聴衆が集まり，世界を牽引する研究者らの関心度が非常に高いことを示していた。個人的な所感は以下である。現在は，学習者コーパス研究の分野も多岐に渡るため，分野によって抱える課題が異なる。本学会も開催期間を数日延長していただければ，各分野に特化したセミナーやワークショップを開催することが可能になり，これまで以上に自分の近い分野に関心がある研究者と知り合い，意見交換できる機会が増えるだろう。

学会に先立ち 10 月 4 日に開催された Pre-conference Workshop では，“LCR at the interfaces”というタイトルで，Granger 氏が Université catholique de Louvain の Centre for English Corpus Linguistics のセンター長を退官するにあたり，氏の功績を称えるものであった。Granger 氏がご自分の研究者及び教育者人生を振り返る講演の後，LCR と SLA（Vyatkina 氏），LCR と対照分析（Hasselgård 氏と Ebeling 氏），LCR と辞書学（Herbst 氏），LCR と自然言語処理（Meurers 氏）が続いた。個人的には，最後の講演での Granger 氏のコメントが非常に印象的であった。学習者コーパスを自然言語処理に積極的に活用していくには，その実現に向け，研究目的に応じた詳細かつ系統立ったアノテーションを施したデータセットを準備するだけでなく，アノテーション・スキーム構築における規準策定やアノテーションの信頼性の向上など様々な課題に留意し，克服すべきであるというものである。

Bolzano/Bozen という地名は，前者がイタリア語名，後者がドイツ語名である。オーストリアとイタリアにまたがるアルプス山脈東部の地域，いわゆる南チロル地方に位置する。ここ Bolzano では，ドイツ語とイタリア語が公用語の他，北部の山岳地帯であるドロミテではラディン語と呼ばれる言語も話されているそうだ。開催場所の EURAC research は，コーパス言語学や Multilingualism の研究も担う研究機関である。筆者は，第 1 回より口頭発表者として本学会に毎回参加させていただいているが，参加者の多くが CEFR の多言語政策を念頭に置いて言語教育を捉える EU 諸国を中心とする研究者や教育者である。イタリア語とドイツ語を巧みに使い分ける人々が暮らす Bolzano という土地は，まさしく CEFR の源流を垣間見るようであり，主催者がかの地を開催都市として選んだ意義を強く感じることとなった。なお，次回の開催場所は，ポーランドのワルシャワということである。

※LCR 2017 については http://lcr2017.eurac.edu/ を参照ください。

写真 1：会場の入り口 1

写真 2：会場の入り口 2
（壁に EURAC research と書いてあります）

写真 3：会場の中庭

写真 4：大ホール（開会式の前）

写真 5：会場の隣にある川辺から望める景色（ドロミテ山塊の一部）

2017 年 12 月 28 日発行

編集・発行　英語コーパス学会
会長　　投野　由紀夫
事務局　　〒157-8511 東京都世田谷区成城 6-1-20
成城大学社会イノベーション学部
石井康毅研究室気付
e-mail: jaecs.hq@gmail.com
URL: http://jaecs.com/