

英語コーパス学会 Newsletter No. 85

Dec 29, 2018

■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

JAECS
Japan Association for English Corpus Studies

<第44回大会報告>

■概要

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)の2日間にわたり東京理科大学(神楽坂キャンパス)にて開催されました。第44回大会では、Stefan Evert先生(Friedrich-Alexander University of Erlangen-Nürnberg, Germany)の講演、井上永幸先生(広島大学)、内田諭先生(九州大学)によるワークショップ、Laurence Anthony先生(早稲田大学)の司会進行によるシンポジウム、さらに14件の研究発表が行われました。前大会に比べ、発表が21件から14件に減りましたが、講演とシンポジウム、研究発表のうち2つのセッションが英語で行われるなど学会の国際化を認識できる大会となりました。

まず、大会初日の午前中には、2件のワークショップが開かれました。最初に、内田諭先生によるワークショップ「言語研究のためのWord2Vec入門」では、自然言語処理の研究で用いられているWord2Vecの紹介が行われました。次に、井上永幸先生によるワークショップ「コーパス研究の作法—語法研究・辞書編集を通して—」では、コーパス研究を始めた初心者が持つべき基本的な考え方や使い方についてお話いただきました。

1日目午後の開会式では、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、開催校である東京理科大学副学長の渡辺一之先生にご挨拶いただきました。

次に、清水眞先生(東京理科大学)の司会のもと総会が行われました。はじめに、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。また、会計の宇佐美裕子先生(東海大学)より、2017年度会計報告及び2018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)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。

午後の研究発表は2室でのパラレルセッションとなりました。第1セッションは柴崎礼士郎先生(明治大学)、第2セッションは水本篤先生(関西大学)の司会のもと、計5件の研究発表が行われました。その後、Laurence Anthony先生の司会進行でシンポジウム「Issues on multi-word units (MWUs) and collocation」が行われました。James Rogers先生(名城大学)、水本篤先生、Stefan Evert先生の3名の先生方からコーパスを用いた熟語やコロケーションの研究方法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。

2日目の午前には、第3セッションは久保田俊彦先生(明治大学)、第4セッションは能登原祥之先生(同志社大学)の司会のもと、計5件の研究発表が行われました。また、午後は、第5セッションは秋山孝信先生(日本大学)、第6セッションは阿部真理子先生(中央大学)の司会のもと、計4件の研究発表が行われました。研究発表に続き、大会の最後に、Stefan Evert先生の講演「Measures of Productivity and Lexical Diversity」が行われました。

第44回大会は2日間を通して、130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。

以下の概要は、発表者・司会者・ワークショップ講師に、大会後に執筆していただいたものです。ただし、講演とシンポジウムについては大会資料に掲載された概要を再掲していません。

■研究発表の概要

第1日第1セッション

[司会] 柴崎礼士郎(明治大学)

「英語史的コーパスによる言語変化速度の測定」

塚本聡（日本大学）

古英語以来、英語には多くの統語的変化が生じている。各種の変化については多くの研究が行われ、変化の開始時期や終了時期、あるいは量的な変化は示されるものの、言語変化に要した時間を横断的に扱った研究は見られない。本研究では、以下に挙げる言語項目の変化に要した時間を観察し、その変化速度にある一定の規則性、法則性がみられるかを研究課題とした。異なる言語資料を使用することによる差異が生じることが無いよう、本研究では Penn-Helsinki Parsed Corpus (ME2, EME, MBE2) および YCOE を使用し、構文解析された統語情報から条件に合致する構造を検索し、その生起数や対立する形態・構造との構成比率の変化を測定することにより変化を観察した。

- (1) 非人称動詞 like の人称動詞化
- (2) 変移自動詞の助動詞選択
- (3) 進行相の確立
- (4) V+-ing 補部の確立
- (5) 属格形および of 迂言形の交替
- (6) 限定詞の義務化
- (7) 二重限定詞
- (8) 二重比較級・最上級

(1) 非人称動詞 like は、1400 から 1600 年にかけて人称動詞化した。(2) be+V-en であった変移動詞の助動詞が have+V-en に変化した現象は 1650 年から 1900 年に変化が完了した。(3) be+ing 構文の主要な動詞が非有界動詞から有界動詞へと拡大した変化は、1650 年には始まり 1900 年には終了した。(4) 動詞補部に-ing が生起する構造は、1625 年から 1900 年にかけて増加した。(5) 屈折語尾の衰退に伴い前置詞の使用が増加した変化は、1100 年には始まり 1300 年には変化が完了した。(6) ほぼ同義となる thereof および of it 構造のうち、thereof は 1475 年頃から急増したが 150 年程度ののちには衰退した。(7) 可算名詞において、a, the, my などの限定詞が必須となる変化は、OE から限定詞の共起率が高まり、700 年かけて現在の共起率に達した。(8) these my words のような限定詞の連鎖は、1450 年頃に使用率が上昇したが、250 年後には衰退した。(9) most easiest の様に迂言形を屈折による比較級・最上級と共に使用する現象は、1375 年頃に使用が増加したが 1550 年頃には非標準用法となった。

定着した(1)~(5)の変化は、ほぼ 200~275 年くらいの時間で変化が完了していること、一方

定着しなかった(6), (8), (9)の変化は、70 年から 125 年程度の増加が継続した後減少に転じている事実から、言語構造の変化に要する時間は 200 年~300 年であったこと、その期間は複数の言語変化に共通していたこと、すなわち変化の速度に関してほぼ一定の類似性があることが確認された。

「アメリカ英語における分離不定詞使用に関する研究：“splitter”に焦点を当てた分析」

福本広光（大阪大学大学院生）

本発表では、アメリカ英語における分離不定詞使用の史的実態について、特に *to* と原形不定詞とを「分離」する要素（副詞）である ‘splitter’ を中心に調査・考察した結果について報告した。本研究では、主として通時コーパスである COHA を用い、splitter となる副詞を以下のように単数語・複数語に分け、分析を行った。

- (1) He seemed to really want to find a way to serve.
(単数語 splitter) [TIME Corpus]
- (2) The states he’s won, the red states, is they’re not likely to all of a sudden turn blue in November.
(複数語 splitter) [COCA]

単数語 splitter を伴う分離不定詞の使用頻度は 19 世紀後半になって一度伸びたが、20 世紀前半に落とし、後半に再度大きく伸びた。変化のメカニズムを探るため、テキストジャンルごとに頻度変遷を調査すると、その背景に 20 世紀前半の Fiction ジャンルでの減少、そして 20 世紀後半の Magazine ジャンルでの増加があることが判明した。（ここでは TIME Corpus も使用）これらの要因として、20 世紀初頭の Fiction ではアメリカ文学の転換期と重なる（川上編, 1999 etc）こと、当時劇や映画のスク립トで比較的多く見られたが、口語的文脈以外において使用が減少したこと、また 20 世紀後半の Magazine でも、使用の幅が広がってきたことを指摘した。複数語 splitter は最大 4 語のものまで見られ、いずれも量的にはわずかながら、口語的なスタイルにおいて使用されていた。

また、splitter の経年的変化を観察するため、COHA の Fiction ジャンルから splitter として用いられている副詞を 50 年分ごとに 4 期間に分け、全て抽出、集計した。結果として、その中には *even, just, really* などの長年にわたって使用されている副詞もあるが、*so, thus* などの、現在頻度を減らしている副詞や *actually, better* などの、現在頻度を増やしている副詞も見られ、時代ごとに特徴的な副詞が存在することが明らかになった。複数語 splitter は、(i) *at once* や *at*

least のような lexicalize されたもの (ii) *more strongly* や *at least passively* のような、一方の副詞が他方の副詞を修飾するものの 2 パターンに分けられることがほとんどで、(i) はどの時代においても固定された表現しか使われないが、(ii) は比較的多様な表現があること、どの時代でも頻度的には (i) のパターンが圧倒的に多いということを示した。

そして、分離不定詞の頻度増加の要因の 1 つと考えられている (Crystal, 1984: 29-30 etc), 分離不定詞の文法的特長について考察した。今回は「ambiguity を解消する」ということについて、splitter が *really, fully* の場合について用例を抽出し、調査を行った。すると、分離不定詞を使わなくても文意が曖昧にならない場合での使用が一定数以上見られた。頻度増加は、文法的特長への注目だけでなく、ルーズな使用が少しずつ増えてきたことも要因であるとの示唆を得たことを報告した。

質疑では、splitter に *not* を用いた場合の変化はどうかということ、splitter を意味的に分類すると時代ごとに変化があるのかということへのご質問や、複数語 splitter も意味的分類が可能ではないかということ、複数語 splitter の 2 パターンを同等に扱わない方が良いのではないかということ、ambiguity の解消のみならず rhythm の観点からも考えてもよいのではということへのご指摘をいただいた。

「コーパスを利用した 18・19 世紀オーストラリア文学作品における自動詞完了形助動詞に関する研究」

守屋輝 (京都大学大学院生)

本研究は、*go, come* 等動作主の移動・変化を表す自動詞の完了形構文において観察される、助動詞 *be/have* の交代現象と、言語使用者の出身や性別といった社会言語学的要因との関係解明を目的に実施した。元来自動詞の完了形には、次の例文のように *be* 動詞が主に使用されていた: Mr. Sandford said, “Hannah what is become of your frock?” (COOEE 2-308)。他動詞の完了形構文で使用される助動詞 *have* が自動詞完了形に対して用いられることが増加し、後期近代英語期(18・19 世紀)には移動・変化を表す自動詞に対しても *be* 完了形の使用頻度を *have* 完了形が上回ったことが、Rydén and Brorström (1987), Kytö (1997) 等多くの先行研究で指摘されてきた。

Have 完了形の増加現象を研究する上で、各植民地社会で新たな言語変種が成立しつつあっ

た後期近代の特性から鑑みて、地域・社会的要因を検証する余地が多分にあると発表者は考えた。本研究では特に、話者の出身地域と完了助動詞選択との関係という側面に焦点を当てた。例えばアイルランド移民は 19 世紀以前オーストラリア入植者の約 20% を占め、イングランド移民に次いで多く (Fritz 2007: 22-27), アイルランド英語由来の統語的特徴の一部がオーストラリア英語 (AusE) 特有の表現として定着したとされる (Burrige and Musgrave 2014: 40-41)。自動詞 *have* 完了形の発展における地域変種の影響を解明する上で、初期オーストラリア社会を対象とした社会言語学的研究は有意義だと考え、本研究に着手した。

本研究では 18-19 世紀のオーストラリアで執筆された文学作品を収録した AustLit 及び Corpus of Oz Early English (COOEE) の 2 種類のコーパスから抜粋した、約 300 万語を対象に分析を行った。Rydén and Brorström (1987) を参考に、19 世紀において *be/have* の交代が確認される出現頻度の高い自動詞 12 種を選び、各自動詞の完了形構文中で *have* が選択される比率を計算して、その結果をテキスト著者の出身地・性別という観点から分析した。

AustLit を用いた通時的分析の結果、*go, come, return* のように他の自動詞より遅れ 19 世紀後半に *have* 使用率が増加しているものと、*improve, recover* のように *be* 使用率が逆に増加傾向にある動詞があることが判明した。また、助動詞 *be/have* の選択と AustLit, COOEE のテキスト著者の出身地との間に弱い相関がある可能性が示唆された。例えば、アイルランド系著者の文章では *be* の使用率が比較的高く、スコットランド系及びオーストラリア生まれの著者は *have* 形が高頻度で現れる傾向が確認された。さらに、性別に基づく分析では同時代のイギリス英語・アメリカ英語に見られる “women writers remain more conservative and systematically prefer the form *be*” (Kytö 1997: 51) という傾向に反し、顕著な男女差は認められなかった。これらの分析結果より、初期 AusE 特有の社会言語学的要因が同地域における自動詞 *have* 完了形の発達に影響した可能性が導かれた。今後も構文やジャンル毎の差異等、より多様な観点からの分析を加え、AusE における完了形助動詞選択の変化に影響した要素の検証に取り組む所存である。

第 1 セッションの概要

柴崎礼士郎 (明治大学)

2018年10月6日大会初日第1セッションでは、英語史に関する3件の発表があった。どの発表でも参加者から多くの質問やコメントがあり、終始和やかに、発表者との建設的な質疑応答が十分に行われる内容であった。以下、発表毎の質疑応答を簡潔にまとめる。

第1発表者の塚本聡氏（日本大学）は、英語史に確認できる9つの構造変化を取り上げ、そのうちの幾つかは200~250年という期間で変化が完了するという調査結果報告であった。この変化の時間幅に質問と関心が寄せられた。例えば、こうした時間幅は言語獲得/習得と関係付けることはできないか、あるいは、構造変化とは別の要因があるのではないか、などの質問があった。生成文法の面から考えると、理論的にはより早くパラメータが切り替わると考えられるが、調査結果からはその裏付けが取れていないとの解答が発表者からあった。また、考察した構造変化は、主節と従属節とで異なる分布を示すのかという質問に関しては、未調査のため今後の課題にしたいとの解答であった。

第2発表者の福本広光氏（大阪大学大学院生）は、アメリカ英語における分離不定詞の変遷について発表した。分析手順として、toと原形動詞を分離するsplitterを1語（e.g. really）から4語（e.g. all of a sudden）までに分けていた。質問の中には、-ly副詞は様態副詞や強意副詞などに分類して、splitterになりやすい副詞の種別が可能かどうか、というものがあつた。また、3語splitterの中にはat least momentarilyのような場合も含まれるため、意味的なユニットとして捉えることは可能かという質問もあつた。前者に関しては分類可能との解答が発表者から得られたが、後者に関しては、時間的制約の中で十分に議論し尽せなかつた。

第3発表者の守家輝氏（京都大学大学院生）は、18-19世紀のオーストラリア英語に見られる自動詞完了形の助動詞選択の調査発表を行った。質問の中には、比較的使用頻度の高いcomeの場合、副詞や補部などの要素を伴うか否かでhave/beの選択に影響がないかというものがあつた。発表者からは、そうした付加要素が無い場合にはbeが選択されやすいという解答があつた。助動詞選択は、総じてhaveになる歴史的傾向が確認できるが、最後までbeとの共起を好むimproveとrecoverにも質問が寄せられた。発表者の解答は、意味的にresultative（結果相）の解釈になる場合にはbeが好まれるという調査結果が紹介された。

セッション全体を通して、聴衆からの質問はどれも的確で、発表者の今後の研究につながる

内容であつた。発表者の対応からは、発表内容以外の背景知識を有し、関連調査に余念がないという印象を受けた。英語史と言う共通点の旗の下、theme sessionとも呼べる整合性のある内容であつた。

第1日第2セッション

[司会] 水本篤（関西大学）

Using Corpora to Examine Lecturing Styles in American and Japanese University Engineering Courses

Judy Noguchi (Kobe Gakuin University)

Kazuko Tojo (Osaka Jogakuin University)

Nilson Kunioshi (Waseda University)

The rapid globalization of society today has accelerated the need for internationalization of higher education in Japan. In 2008, the Japanese government announced the “300000 Foreign Students Plan” Campaign (Ministry of Foreign Affairs, 2010) and as of May 1, 2017, the number of international students studying in Japan had reached 267,042, with 188,384 students enrolled in institutions of higher education (Japan Student JASSO, 2017). This has led to an increase in the demand for university degree courses which are offered using English as the medium of instruction. With an interest in aiding Japanese instructors faced with delivering lectures in their disciplines in English, especially those in science and engineering, we built OnCAL (Online Corpus of Academic Lectures), a corpus of university lectures for science and engineering courses. We examined these lectures and identified their pedagogical functions (e.g., science chronology, cause/effect, thought experiment, question) and useful expressions to express these functions. The OnCAL concordancing interface allows access to 430 science and engineering lectures given at MIT OpenCourseWare (MIT OCW, <http://ocw.mit.edu/index.htm>) and Stanford Engineering Everywhere (SEE, <http://see.stanford.edu/>). This work led us to wonder about how the functions that we had uncovered, such as the asking of questions and the proposal of thought experiments to initiate consideration of the lecture contents, would be delivered in comparable lectures at Japanese universities. We therefore began construction of a corpus of Japanese lectures given at four major national and private universities in similar disciplines. At present, there are 104 Japanese lectures and comparing them with the

American lectures revealed marked differences in the lecturing styles. In this report, we focused mainly on questions and their functions in the lectures. In general, the lectures given at American universities included many question-word phrases to elicit student thinking with frequent uses of personal pronouns such as *you* and *we*, indicating more interaction during the class. Question-word phrases with *what* (17.1 average per lecture), *how* (6.97), and *why* (2.44) were frequently used, in contrast to the Japanese lectures where 何, なん (*what*) was used only 2.8 times on average per lecture, なん(て), どう, どのように (*how*) only 2.8 times and 何(で), なん(で), どう(して), なぜ (*why*) only 0.6 times. Another striking difference noted was the much more ample usage of modal auxiliary verbs in the English lectures. Overall, the American lectures, which included many student-initiated questions, tended to be presented in a conversational style with a strong audience orientation. This was in contrast to the Japanese lectures which were given in a more formal style suggesting the presence of a top-down authority and an emphasis on content dissemination. Our findings led us to conclude that English-medium instruction may not be successful if a lecture originally intended for a Japanese university audience was just simply delivered in the English medium. What is essential for successful instruction in a “globalized” classroom with students from different educational backgrounds is an awareness of differences in lecturing styles and the structuring of lectures in order to reach students who may have different expectations with respect to classroom instruction.

Collaborative Texts under a Stylometric Microscope:
Investigating Texts of Mixed Authorship
Tomoji Tabata (Osaka University)

(大会後に発表者から研究発表報告が提出されませんでしたので、本発表の報告はございません。大会資料の要旨をご覧ください。)

第2セッションの概要

水本篤 (関西大学)

The first presentation was cancelled. The second presentation was “Using Corpora to Examine Lecturing Styles in American and Japanese University Engineering Courses” by Judy Noguchi (Kobe Gakuin University). In the Q&A time, a question was raised to confirm if the questions in the

classroom included those from students. As the target of the current study was questions asked by teachers, the answer was no. Professor Noguchi further commented that when she first embarked on investigating the question types using the OnCAL corpus, it turned out that the differences were far more dramatic. Another question was if the OnCAL corpus has an annotation for the class size. Professor Noguchi replied that there is no information about the class size in the OnCAL corpus.

The third presentation was “Collaborative Texts under a Stylometric Microscope: Investigating Texts of Mixed Authorship” by Tomoji Tabata (Osaka University). During Q&A, the first question was if the technique used in this study could be applied to identifying anonymous authors who send letters to newspaper companies. Professor Tabata answered that it would be possible, but the shorter the text, the harder the detection of the author would be. The second question concerned how he extracted the “author makers” from the corpora. He described that the effect sizes of Mann-Whitney’s U test were used for obtaining the author makers. Professor Tabata also mentioned that the same exact result was observed using the bootstrapping method. The third question posed was, if one author took over the other in the middle of writing, the subsequent author may not have changed the writing style out of consideration to the author who initially wrote it. Professor Tabata replied that it might have been the case.

The two presentations were thought-provoking in that they used and analyzed corpora in a very innovative way, and they also investigated their respective research questions both quantitatively and qualitatively.

第2日第3セッション

[司会] 久保田俊彦 (明治大学)

「Charles Dickens の *The Mystery of Edwin Drood* の, “Thomas Power James”による続編の『著者推定』—テキスト全体・地の文・発話の3種類の語彙による推定の比較・評価」

後藤克己 (中部大学大学院生)

米国人 Thomas Power James (以下, James) は, Dickens の遺作となった *The Mystery of Edwin Drood* (以下, *ED*) に続編を加えた「完全版」(Dickens, [James], 1873)を公表し, その続編を“By the Spirit Pen of Charles Dickens, through a Medium.”とアピールしている。この続編につ

いて W.H.B (1874), Gadd (1905), Doyle (1927), Wolkomir (1973)などが精読による批評を加えているが、このアピールの妥当性についての結論は得られていない。

本研究では多次元尺度法およびクラスター分析を使用し、語彙嗜好の観点から続編の著者が「Dickens の霊」と言い得るかどうかの推定を試みた。この推定は、続編が Dickens の霊に導かれてテキスト化されたものであれば Dickens の文体特徴をもつはず、との考えによる。なお著者推定には、一般に地の文に生起する語彙頻度が多く用いられる(Hoover, 2001, 2002)が、発話テキストの除去には、1 重引用符とアポストロフィの区別、引用符を欠いた自由直接発話の抽出などが必要となり大きな負担となる。そこでテキスト全体/地の文/発話のそれぞれの語彙で分析し、分析語彙を地の文に限定することの効果の評価した。

コーパスには ED と続編に、参照用として Dickens の *Our Mutual Friend* (以下, OMF) を加えた 3 作品を用いた。これらを地の文と発話に分離し、作品毎に 4~12 のセクションに分け、語彙の Lemma による生起数を抽出した後、Hoover (2004)を参考に 1 作品に 70%以上偏在する語、人称代名詞および固有名詞を除外し、テキスト全体・地の文・発話の 3 種類の語彙について、それぞれ生起頻度上位 100 語~1,000 語の 6 つの分析用データを準備した。

分析の結果 ED・OMF のセクションは 1 つのクラスターを形成するものの、続編のそれは別クラスターを形成しており語彙嗜好が異なると考えられることから、続編を Dickens の「遺作」とする James のアピールは疑わしいとの結論を得た。なお、多次元尺度法のプロットおよびクラスター分析の樹形図において、続編と ED・OMF との距離は、地の文語彙の場合が最も大きく、地の文語彙による著者推定性能の高さが確認された。また、上位 1,000 語を 100 語毎にセグメント化して分析し、上位 200 語程度までの高頻出語彙が著者推定への高い影響度を持つことを示した。さらにそれらが平易な日常語であることから、その多くは著作にあたって意識的に語彙選択される語ではないと推定し、そのような語彙で Dickens の 2 作品と続編の語彙嗜好の差異が確認されたことは、James のアピールへの疑問をより深めるものと結論づけた。

「トピックモデルを用いた Agatha Christie 作品へのアプローチ」

土村成美 (大阪大学大学院生)

本研究では、イギリスのミステリー作家 Agatha Christie の作品に関して、同時代作家と比較して統計的手法を用いた特徴の分析を行うことを目的とする。比較対象として、Christie と同時代に活躍した女性ミステリー作家、Dorothy Sayers, Margery Allingham, Ngaio Marsh の作品を用いる。これら 4 名の女性作家は第一次世界大戦と第二次世界大戦の戦間期を中心に活躍しており、「イギリスミステリーの 4 大女王」と呼ばれているためである(Joannou (Eds.), 2013)。

分析に用いるデータは以下の通りである。Christie 作品 66 作品(4,183,485 語), Sayers 作品 11 作品(1,115,019 語), Allingham 作品 20 作品(1,534,462 語), Marsh 作品 33 作品(2,510,391 語)である。分析対象作品は長編作品に限定している。作品ごとに総語数が大きく異なるため、各作品のテキストファイルを 2000 語単位に分割した。この処理を経た合計 4661 ファイルを用いて分析を行った。

分析手法として、機械学習の一種であり、確率論的アルゴリズムに基づいてトピック(話題, テーマ)の抽出を行うトピックモデルを用い、分析を行う。本研究では Blei et al. (2003)によって提唱された潜在的ディリクレ配分法を用いたトピックモデルを行う。トピックモデルを実行するにあたり、マサチューセッツ大学で開発された自然言語処理ツールキット MALLET (Machine Learning for Language Toolkit)バージョン 2.0.7 を使用した。この分析を通して、Christie 作品において同時代作家と比べて特徴的なトピックを抽出することが本研究の目的である。

Christie 作品と特に関連性が強いトピックを確認すると、第一に挙げられるのは、縮約表現を中心とした口語的表現から成るトピックである。このトピックには *I'm, that's, didn't, I've, you're* のような語が含まれている。Christie の作品は登場人物の会話を中心として物語が展開されるものが多く、その特徴が反映された結果であることが考えられる。また、このトピックと関連する作品は Christie の初期の作品よりも、晩年の作品の方が多くなっている。Christie は 1952 年に腕を骨折し、従来までの手書きの原稿をタイプライターで打つという執筆方法から、ディクタホンを使用した執筆へと、執筆スタイルが変化している(Le et al., 2011)。この執筆方法の変化も Christie 作品がこのトピックと関連性が強い要因となっていると考えられる。

全作家に対してミステリー作品のみでコーパ

スを構築しているため、*case, murder, death, evidence, police* などの語が含まれる犯罪捜査に関するトピックも生成されており、このトピックに関しても Christie 作品との関連性が強く出ている。Christie が直接的に犯罪を表現する語を多く使用していることが反映されている可能性が高いと言える。また、トピック数を増やすと Christie 作品の探偵 Poirot の言動を表す語から成ると推察されるトピックも現れることが確認された。同ジャンル作品を対象としたトピックモデルで、どの程度作家間の特徴を識別することが可能であるかに関して可能性を検討する。

第3セッションの概要

久保田俊彦（明治大学）

第3セッションでは計量的著者推定、計量文体論に関する2件の発表がなされた（第二発表がキャンセルされたため）。

第一発表者の後藤克己氏（中部大学大学院生）の発表は「Dickens の霊によって書かれた」という触れ込みで出版された「Dickens 未完遺作の続編」の真贋を問うユニークな著者推定研究であった。高頻度語彙を変数とした多次元尺度法により、この続編は他の Dickens 作品とは異なる性質を持つことが示される。質疑応答では、Dickens の霊の真筆性を肯定的にとらえた批評についての質問のほか、使用データの拡大や、語彙難度分析に使用したベース辞書の変更などが推薦された。

第三発表者の土村成美氏（大阪大学大学院生）の発表はトピックモデル (LDA) を使用した Christie の計量文体研究であった。ヒートマップ、ワードクラウド、ネットワークグラフなどによって視覚化された分析結果をもとにトピック毎の解釈が示され、この解釈と先行研究との関係にも多くの言及があった。質疑応答では、トピックモデルで得られる「トピック」の意味、探索的調査におけるトピックモデルのさらなる応用の可能性などについてやりとりがあった。

両発表とも、分析対象への理解に基づいて、行なわれるべき分析が丁寧に設計されている点に好感が持てた。すなわち計量文体研究とは作品データと分析法（ソフトウェア）の単なる掛け算ではないことが示されていたと言えるだろう。

第2日第4セッション

[司会] 能登原祥之（同志社大学）

「小学生のための英語 DDL 支援サイトの開発に向けた作例参照用コーパスの構築と検索ツールの開発」

西垣知佳子（千葉大学）
赤瀬川史朗（Lago 言語研究所）
中條清美（日本大学）

本発表では、小学校英語に DDL (Data-Driven Learning: データ駆動型学習) を導入するための「小学生のための英語 DDL 支援サイト」の開発について報告した。発表者らが開発した DDL 支援サイトは、小学生の英語力レベルにあった英文を集めたセンテンスコーパスを教材として用い、小学生がそれを検索して英文のルールを学ぶものである。今回の開発では、大学生用 DDL サイト『SCoRE』(中條他, 2018) の手順にならった (図1)。

発表では、はじめに、開発した小学生用 DDL 学習支援サイトのプロトタイプを実演し、続いてその開発手順を、主に「参照用コーパス」の

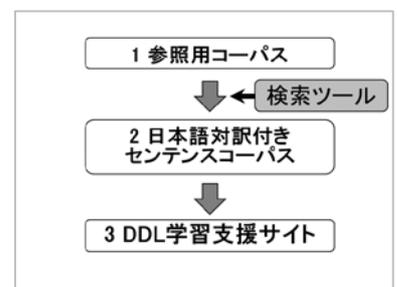


図1 支援サイトの開発手順

構築 (図1の1) ならびに「検索ツール」の開発 (図1の網かけ) を中心にして報告した。「参照用コーパス」は、「日本語対訳付きセンテンスコーパス」作成 (図1の2) の際に参照するための入門期・初級者向けの例文を集積したコーパスで、2,500万語規模の入門期英語を収集した。ソースとなるデータは、アメリカの教科書、東アジア諸国の検定教科書、Graded Readers、子ども向けのニュースなどのインターネット上の入門期英語コンテンツなどである。集めたデータは、センテンスに区切り、Stanford POS Tagger を用いて品詞タグ付けを行った。

次に、小学校英語における DDL 実践では、入門期学習者の英語のルールに対する気づきを引き出すために、精選された英文の用例の提示が不可欠である。そのため提示する英文の長さ、語彙や文法のレベル、内容やトピックなどを学習者のレベルに合わせて調整する必要がある。そこで、そのような用例を、「参照用コーパス」から効率よく検索するための「検索ツール」を開発した。この「検索ツール」では、(1) 検索結果をセンテンスの短いものから順に表示できる、(2) 検索するセンテンスの長さを語数

で指定して検索できる, (3) CQL(Corpus Query Language)を使ってレマや品詞を指定した検索ができるので, 学習ターゲットの文構造を設定して検索できる。これらを可能にするために, 「参照用コーパス」にはセンテンス長の情報を付与した。また, センテンス長とセンテンスの文字列をソートキーとしてコーパスの全用例を並べ替え, 重複するセンテンスを排除した。さらに, センテンス長が 1 語から 100 語までの用例のサブセット(約 197 万センテンス, 約 2,234 万語)を作成し, 作例参照の用途により適した形に再編集した。

最後に, これらの開発によって, 小学生用 DDL サイトで提示する「日本語対訳付きセンテンスコーパス」の作成作業の効率が向上したこと, また, これらを利用して, 現在, 一般公開を目指して小学校英語のための DDL 支援サイトの改定を行っていることを報告し, 発表を終えた。

「英日バイリンガルエッセイコーパスに見るコロケーションの比喩的な意味拡張」

鎌倉義士 (愛知大学)

第二言語学習者の習熟度を測る指標としてメタファーを含む比喩表現の使用が注目されている。metaphoric competence (Littlemore, 2001) の概念において, 目標言語での比喩表現の理解や運用の度合が学習者の言語の熟達度と関連すると言われる (Littlemore, Krennmayr, Turner & Turner, 2014)。本研究では日本人英語学習者の比喩表現使用の分析に, 英語と日本語の両言語でのエッセイをデータとするバイリンガルコーパスを使用した。なぜなら, 学習者が英語で表現できなかった比喩も対訳の日本語作文でその内容を確認することが可能であるからである。特にコロケーションに注目し日本人学習者の英作文に確認される基本的な意味から比喩的な意味への拡張について分析を行った。例えば, big, serious, significant, important など重大さを示す形容詞が problem(s)と共起する頻度が高い一方, complicated, complex という困難さを示す形容詞や basic, central など基礎的な意味を指す形容詞は日本人学習者の英作文ではネイティブよりも頻繁に使用されていない。このように日本人英語学習者のコロケーションの特徴を problem(s)との共起に限定して本研究では分析を試みた。

研究課題は以下の 3 つとなる。1) problem(s)と共起する形容詞・名詞・前置詞の分析から日本人英語学習者に特徴的な problem(s)の意味拡

張ネットワークが存在し, 記述することは可能か。2) problem(s)の意味拡張ネットワークは単純に対訳語となる「問題」とその共起する日本語の語彙の組み合わせとは異なり, 日本人学習者の英語独自の意味拡張が観察されるか。3) TOEIC スコアを基準にエッセイ作成者のグループを三つに分け, その上位と下位グループの間に意味拡張の使用頻度と英語の熟達度が関係するのか。

研究課題として述べた日本人英語学習者特有の意味拡張ネットワークは図 1 の「形容詞+problem(s)」のネットワークに加え, 表 2「名詞+problem(s)」と表 5「problem(s)+前置詞」からもその傾向が確認された。そのコロケーションから見る比喩的な意味拡張は, 単に母語からの影響だけではなく, 英語学習を通じて育まれた独自の発展である。

発表後の質疑応答では, 本研究にて提示された図に加えて, 母語話者の使用を分析し同様に図を作成することで視覚的に比較が可能となる提案があった。そして, コロケーションの意味拡張が示す内容とその意図を明確にすべきという意見があった。今後の研究の参考としたい。

本研究の目的は英語母語話者が使用するコロケーションと比較し, 第二言語学習者である日本人学生のコロケーション使用の狭小さや貧弱さを指摘することではない。むしろ, 日本人に好まれる英語でのコロケーションを分析することで日本人英語学習者の比喩表現を記述し, 日本人らしい英語を解釈可能なコロケーションと認めた上で更なる上達を促し, 未発達の意味拡張の方向性を提示するのが目的である。日本人英語学習者は比喩表現の使用に積極的ではない傾向があるが, 発信するための英語の上達には比喩表現の学習は不可欠である。

「英語学習者コーパス構築のためのタスク設計：特定の文法項目抽出に向けて」

工藤洋路 (玉川大学)

内田諭 (九州大学)

英語学習者コーパスから得られる学習者の文法の使用状況は, 言語テストの開発や教材作成, そして効果的な指導法の確立などに対して, 有益な情報を提供することができるが, 各文法項目と学習者のレベルの関係を調べるためには, コーパスに蓄積されている学習者の言語データがどのようなタスクを用いて抽出されたかという情報が必要になる。例えば, 同じ学習者でも, あるタスクの下で行われたライティングでは, ある特定の文法項目を使用するが, 別

のタスクの下では、その文法項目を使用しないという現象が起きることは容易に想像がつく。つまり、学習者が特定の文法項目を習得していても、対象タスクではそれを使う必然性が生じず、その項目を使用しないということは起こり得るということであり、タスクと特定の文法項目の出現のしやすさについて検証が必要である。そこで、本研究では、「特定の文法項目を学習者から必然的に引き出すためのライティングタスクの要件は何か」をリサーチ・クエスチョンに設定した上で、文法項目抽出を目的とした複数のライティングタスクを設計し、それぞれのタスクデザインが学習者の文法使用にどの程度影響を与えるかというリサーチを行った。特定の文法項目が使用される確率が高いタスクを開発することができれば、学習者の文法習得状況を効果的に測定することが可能となり、また文法項目を軸とした学習者コーパスの構築が可能となる。

本研究では、高校の「英語表現 I」の教科書のライティング活動を用いて、2つのグループの参加者に、それぞれ異なる設計のライティングタスクを実施した。対象となる文法項目は受動態、現在完了、不定詞の名詞的用法（形式主語の it）で、1つ目の学習者グループには、(1)教科書の活動からモデル文を取り除いたタスク（指示文のみ）と、(2)教科書の活動（指示文あり、モデル文あり）をそのまま利用したタスクの2つを実施した。2つ目のグループには、(3)教科書の活動からモデル文を取り除いたタスク（指示文のみ=(1)）と、(4)教科書の活動からモデル文を取り除き、指示文を詳細に書き直したタスクの2つを実施した（指示文改、モデル文なし）。その結果、それぞれのタスクで使用することが想定される文法項目（受動態、現在完了、不定詞の名詞的用法）を使用した学習者の割合は、(1)よりも(2)のタスクにおいて、概して約3割程度高く、特定の文法項目抽出において、モデル文を提示することの有効性が示された。同様に、(3)よりも(4)のタスクにおいて、想定される文法項目（同上）を使用した学習者の割合が高くなり、特に現在完了については約35%から89%へと大きく上昇した。これらの結果から、モデル文の有無や指示文の詳細さなど、タスクの設計方法によって、テーマやトピックが同じライティングタスクであっても、学習者が特定の文法項目を使用できる確率が変わってくるということが明らかとなった。

第4セッションの概要

能登原祥之（同志社大学）

第4セッションでは、DDL支援サイトに関する発表が1件、学習者コーパスに関する発表が2件あった。質疑応答では、特に、コーパスデータを利用する学習環境のあり方、学習者データの分析と解釈のあり方、学習者データを収集するタスクのあり方、などについて活発に意見が交わされた。

まず、第一発表者の西垣知佳子（千葉大学）・赤瀬川史朗（Lago言語研究所）・中條清美（日本大学）の各氏からは、小学生を対象とするDDL支援サイト試作版（CAN-DOで検索可）と検索ツールBEC（Beginners English Corpus）Search（CQLによる検索可）の開発プロセスが紹介された。発表後には、支援サイトやBEC Searchの開発プロセスをどの程度公開予定か、表現したいことに近い和文を探すと適切な英文が見つかるというインターフェイスの方が検索も圧倒的に速く使いやすいのではないかと、などの質問が出た。

次に、第二発表者の鎌倉義士氏（愛知大学）からは、関西大学バイリンガルエッセイコーパスを利用し、日本人英語学習者（大学生）のコロケーション知識（形容詞+problem(s)の場合）と形容詞に見られる比喩的意味ネットワークの拡張に関する研究が報告された。発表後には、英語母語話者の意味ネットワークとの比較は考えていないのか、MI指標とt指標の使い分け方に意図があるのか、学習者の熟達度レベルを判別するTOEICスコアに根拠はあるのか、判別分析は何を基準に分けたのか、など分析法について確認する質問が出た。また、Tyler & Evansのネットワークモデルを参考にする際、形容詞のみのネットワークと考えるのか、形容詞+problem(s)コロケーション全体のネットワークと考えるのか、など学習者の語彙の意味ネットワークの解釈に関する質問が出た。

最後に、第三発表者 工藤洋路（玉川大学）・内田諭（九州大学）各氏からは、あるライティングタスクがCEFRレベルの基準特性とされる文法項目（現在完了、受動態、不定詞の3つの場合）を自然に抽出できるタスクかについて、(1)Vision Questの指示文の書き換えの有無を比較する場合（大学生対象）と(2)Vision Questのモデル文の有無を比較する場合（高校生対象）の2つの調査結果について報告された。発表後には、ジャンル、文体、トピックなどが産出（学習者の英作文）に影響を与えていると思うがどうか、タスクは文法項目を抽出するものではないとする考え方がどうか、現在完了が本当にA2レベルの基準特性か、産出できる

人は産出する（産出できない人は産出しない）
タスクが妥当性の高いタスクだと思うがどう
か、などの質問が出た。

本セッションを通して、コーパスデータの性
質とその背景にある世界をどのように捉え、
データから見えるもの（見えないもの）をどう
解釈し、それをどのように教育で扱うべきか、
など、DDL 研究や学習者コーパス研究を丁寧
に進めていく上で考えるべきポイントについて
さまざまな角度から考えさせられた。

第2日第5セッション

[司会] 秋山孝信（日本大学）

「アカデミックライティングにおけるヘッジの
活用：研究論文における Discussion のコーパ
ス分析」

中谷安男（法政大学）

（大会後に発表者から研究発表報告が提出され
ませんでしたので、本発表の報告はごさいませ
ん。大会資料の要旨をご覧ください。）

「強意語的機能を持つ罵倒語の進化特性につい
て」

新井洋一（中央大学）

この発表では、まず主要な英語の罵倒語
(expletives or swearwords)を5種類に分類し、今
回取り上げた強意語的機能を持つ damn,
fucking, bloody が、それぞれ異なる種類に属す
ることを説明したあと、OED3 (Online)の初例か
ら、bloody (1676), damn (1787), fucking (1893)の
順に、ほぼ100年間隔で、新たな発達を遂げて
来ていることを確認した。また高増(2000),
Hughes (2006), McEnergy (2006), Ljung (2011)など
の既存研究を前提に、これらの共通の特徴を以
下の4点にまとめた。

(1) ①下層階級から発達したこと

②当初の印刷物では、部分的省略、ダッ
シュやアステリスクによる置き換えなどの伏せ
字語であったこと

③語頭に共通の代表的な子音を持つこと: /b/
(bastard, bitch, etc), /d/ (darn, devil, etc.), /f/
(footling, frigging, etc)

④語・句中挿入(infixing / tmesis)が可能なこ
と: abso-**bloody**-lutely, not **bloody** likely (B. Shaw,
Pygmalion), etc.

次に新井(2011)に倣って、快・不快の意味
を表す意味素性として [±PLEASANT] (略して
[±P]) を導入し、BNC に含まれる罵倒語との
共起語を意味特性に基づいて整理し、①「共起

名詞が不快な意味素性[-P]を持つ場合 (a
bloody nuisance, a **fucking rain forest**, a **damn fool**)
が基本特性であること」②「快でも不快でもな
い中立的な意味素性[±P]の場合 (that **bloody**
bird, this **fucking city**, the **damn thing**) は、共起
名詞に対して不快な意味合いを付帯する意味的
韻律 (semantic prosody) 現象が感知できるこ
と」を明らかにした。さらにこれらが副詞とし
て機能する場合は、③「快適な意味素性[+P]を
持つ形容詞とも共起する場合 (**bloody amazing**,
fucking brilliant, **damn lucky**) があること」、④
「動詞を修飾する場合 (**bloody come**, **fucking kill**,
damn want) があること」を解説するとともに、
このような意味特性の観点からの研究が、
上述のいずれの先行研究にも見られないことを
指摘した。この後者の③④の例によって、BNC
のデータが生まれた1990年代前半にすでに、
この3つの罵倒語の意味の漂白化 (bleaching)
が進み、快・不快の意味の区別を要求しない、
中立的な強意語として機能し始めていると推察
できる。そして、取り上げた3種類の罵倒語の
共通の機能的進化として、次のような品詞転換
(conversion)プロセスを提示した。

(2) adj. (attributive: pre-noun) => adv. (intensifier:
pre-adjective) => adv. (intensifier: pre-verb)

後半では、BNC から約30年近くになる2018
年現在の実態を調査できる大規模コーパスとし
て、NOW corpus (<https://corpus.byu.edu/now/>) を
取り上げ、その規模、主要な特徴、長所・短所
を紹介したあと、今回の3つの罵倒語の共起語
の特性について報告した。NOW コーパスにお
いても、[+P]素性を持つ形容詞との共起例は見
られ、とりわけ、⑤「damn と共起する[+P]形容
詞の種類が格段に増えていること」、また ⑥
「damn と共起する動詞として be 動詞が顕著で
あること」などを明らかにした。

第5セッションの概要

秋山孝信（日本大学）

第5セッションでは、コーパスを用いた英語
学研究に関する二件の発表がなされ、学術論文
の Discussion における効果的なヘッジの活用法
と、英語罵倒語における統語的・意味的特性と
その進化傾向についての研究結果が報告され
た。

第一発表者の中谷安男氏（法政大学）は、氏
の研究チームが作成した学術論文コーパス（自
然科学、社会科学（経済学・経営学）、人文科
学（応用言語学）の三分野において、それぞれ
インパクトファクターの高い代表的な学術誌を

二誌ずつ選び、そこから 102 本の研究論文を選び構築した、総語数 105 万 2 千語からなるコーパス) を分析することによって、より説得力のある Discussion セクションを書くことを目的としたヘッジ表現の用いられ方が示された。特に、先行研究ではあまり検証されてこなかった、Discussion における 3 つのムーヴ (情報の配置) における特徴的なヘッジの活用が明らかとなった。具体的には、*suggest*, *likely*, *would*, *may/might* といったヘッジが Discussion では顕著に観察されることを指摘し、それぞれの用いられ方の傾向が具体例と共に解説された。将来的には、研究者・学生が本研究結果を英語論文作成に応用することが大いに期待できる。質疑応答では、(1)コーパスを Discussion のセクションとそれ以外のセクションに分けて分析が行われていたが、どのような基準でそれらを分けていたのか、(2)今回の分析では、語のレベルのみを対象としているのか、それともフレーズのレベル(to some extent, as far as I know など)も分析対象としているのか、さらに、(3)特定のヘッジ表現とその特徴的な用法 (例えば、*may* や *might* は仮説に反する結果を説明する際に使われる傾向がある等) を、データを用いて可視的に示すことは可能であるのか、といった質問がなされた。

第二発表者の新井洋一氏 (中央大学) は、おもに形容詞や副詞の機能を持つ英語の罵倒語である *bloody*, *fucking*, *damn* を研究対象とし、OED, BNC, NOW corpus などから抽出したデータに基づいて、これらの統語的・意味的特性を明らかにした。(BYU-corpora の中では、NOW corpus から最も多くの罵倒語の実例が検索された。) 先行研究では、罵倒語が形容詞として用いられる場合、不快な意味を持つ名詞 (nuisance など) と共起し、その不快さの度合いを強調する強意語的機能を持つことが指摘されてきたが、新井氏はコーパスの分析により、上述三つの罵倒語が快適な (あるいは肯定的な) 意味を持つ形容詞や副詞と共起することがあるという、近年の際立った傾向についての分析結果を示した。例えば、NOW corpus の分析では、*bloody* が *brilliant*, *fantastic*, *good* といった [+pleasant] な意味素性を持つ形容詞を修飾する事例があることが判明した。さらに *do*, *hope*, *love*, *wait* といった動詞を修飾する例も提示され、罵倒語の大変興味深い統語的・意味的特性と進化傾向が説明された。質疑応答では、研究対象とした 3 つの罵倒語は、OED の historical thesaurus において扱いはあるのか、またなぜ *bloody* は罵倒語として用いられるようになった

のかといった質問が寄せられた。さらに新井氏により、イギリスのテレビドラマにおいて実際に罵倒語が修飾語として用いられている場面の動画が上映された。

第 5 セッションは全体として、コーパスから得られた実例一つ一つを精緻に検討することによって、大変興味深い新たな言語事実を提示し、これまで明らかになっていなかった英語の用法に新たな知見をもたらしていた。コーパスを用いた実証的英語学研究は、分析対象の幅が広く、研究結果がより信頼に値することを改めて感じさせる発表であった。

第 2 日第 6 セッション

[司会] 阿部真理子 (中央大学)

Reliability and Replicability of Annotation Schemes for Learner Corpora

Aika Miura (Tokyo University of Agriculture)

The aim of this study was to compare the inter-annotator agreements for three annotation schemes for a spoken learner corpus, the NICT JLE Corpus. The author developed the following multi-layered annotation schemes and conducted the initial annotations manually: (i) identification of learners' requestive speech acts, (ii) labeling of the functions of learners' utterances, and (iii) assignment of their degree of grammatical accuracy and acceptability in the utterances. In order to examine whether the annotation schemes were "reliable" and "replicable," and to conduct analyses in a "transparent" manner (Fuoli and Hommerberg, 2015, p. 316), the author reports the obtained agreement measure of Krippendorff's *alpha* for each annotation scheme (Artstein and Poesio, 2008; Krippendorff, 2004; Geertzen, 2012).

In the present study, the author examined learner productions during shopping role plays from the NICT JLE Corpus, which contain 68, 114, and 66 files of utterances at the Common European Framework of Reference for Languages (CEFR) levels of A1, A2, and B1, respectively. In the first scheme, the author identified pragmalinguistic features of requests, drawing on the Cross-Cultural Study of Speech Act Realization Patterns (CCSARP) coding scheme (Blum-Kulka, House, and Kasper, 1989): desire verbs (e.g., *want*) and imperatives were categorized as direct strategy, while ability/permission modals (e.g., *can*) and suggestory

(e.g., *how about*) were grouped as conventionally indirect strategies. The second and third schemes were involved with classification of the function of every utterance. The utterances were divided into two major functions: “dealing with transaction” (including subcategories such as “expressing their intention of purchase” and “expressing or asking about the item”) and “communication for transaction” (including “explaining the background,” “requesting an action,” and “confirming”). The degree of grammatical accuracy and discoursal acceptability was also assigned to each utterance: “high” indicated that the utterance was grammatically accurate and acceptable in terms of discourse; “low” was further categorized into (i) coherent (i.e., discoursally coherent but slightly ungrammatical), (ii) slightly incoherent (i.e., semantically inferable but grammatically unacceptable), (iii) incoherent (either discoursally unacceptable due to ungrammatical features or structurally and semantically acceptable but completely incoherent in terms of discourse), and (iv) featuring the use of Japanese. Table 1 shows the total numbers and ratios of the annotated segments.

Table 1

Total numbers and ratios of the annotated segments

The CEFR Level	A1	A2	B1
Total segments for the 1 st scheme	597	1,170	412
Total segments for the 2 nd and 3 rd schemes	893	1,911	1,159
The ratios of the functions			
Dealing with transaction	59.02%	55.03%	0.65%
Communication for transaction	40.8%	44.97%	99.35%
The ratios of the high and low segments			
High	52.18%	54.98%	66.82%
Low			
Coherent	41.72%	41.74%	31.88%
Slightly incoherent	3.68%	2.13%	1.11%
Incoherent	2.18%	1.09%	0.19%
Japanese	0.23%	0.05%	0%

Searching for grammatical items as criterial features of CEFR levels in spoken and written learner corpora: Using the CEFR-J Grammar Profile

Yukio Tono (Tokyo University of Foreign Studies)
Yasutake Ishii (Seijo University)

There is a growing interest in profiling L2 learners’ proficiency levels based on the CEFR, and research projects such as *English Grammar Profile* (2015), *Global Scale of English* (2015), and the CEFR-J (Tono, 2013) seek to identify so-called “criterial features” for distinguishing one CEFR(-J) level from the others.

This study investigated what grammatical items can serve as criterial features of the CEFR(-J) levels to evaluate the English utterances and writings by Japanese EFL learners, and whether two different modes of production, i.e. spoken and written, need different criterial features. Two learner corpora, the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004) and the JEFLL Corpus (Tono, 2007) were used for the analysis of L2 spoken and written production respectively. All data were re-classified according to the CEFR(-J) level of each spoken/written production. Using the inventory of grammatical items developed for the CEFR-J Grammar Profile (Ishii & Minn, 2015), we obtained the relative frequency of grammatical items in each participant’s data. The items used in this study were selected based on their frequency in the whole data in each corpus, whereby 124 items were chosen for the NICT JLE and 196 for the JEFLL.

Random Forest implemented in the “randomForest” package in R was used for binary classification of different levels or level groups (e.g. A1 vs A2, A1 vs non-A1, A2 vs B1.1, and A vs B) in order to search for discriminating features useful for classification. The reason why we tried to separate the data into two level groups, not into all four or five levels present in our data sets, is the rather low accuracy rates we got for multi-level classification in our pilot analysis. The data from odd line numbers were used to train our model while the remaining data was used to test the model’s accuracy.

The results show that in the JEFLL, accuracy rates for telling A1 from other levels is lower than those for discriminating more advanced levels (A1-A2 79%; A1-nonA1 82%; A2-B 93%; A2-B1 93%), whereas in the NICT JLE the opposite relationship is discerned and we can more accurately distinguish A1 levels (A1-A2 92%; A1-nonA1 94%; A2-B 75%; A2-B1.1 71%). (All the percentages given above are approximate average figures.)

Relative weights of grammatical items as predictors were evaluated by the mean decrease in

Gini index. Variables which are effective in distinguishing A1 and A2 levels in the JEFLL include prepositions, coordinating conjunctions, the definite article *the*, and to-infinitives, while in the NICT JLE, *have to* (affirmative), *can* (affirmative), the present tense of lexical verbs (affirmative), and personal pronouns *me/us/him/her/them* are measured to have great contribution to the distinction between A1 and A2.

There seems to be a big difference between the two different learner corpora in terms of what grammatical items count for discriminating the levels and how effective they are. The difference seems to be partly due to the different tasks in the two corpora, but it should be noted that apparently different modes (i.e. spoken vs. written) require different criterial features. In our presentation at the conference, we discussed in detail learners' use of grammatical items as possible candidates of criterial features for their CEFR(-J) levels, together with some methodological and pedagogical implications.

第6セッションの概要

阿部真理子 (中央大学)

The first presenter in this session, Aika Miura (Tokyo University of Agriculture), has studied reliability and replicability of annotation schemes for learner corpora. She has compared the inter-annotator agreements for three annotation schemes and developed multi-layered annotation schemes. In order to accomplish her aim, she has documented manuals, provided annotation training to an external annotator, conducted a random check of annotated segments, and requested the annotator to replicate the annotations of 12 files for each annotation scheme without referencing the manuals. The discrepancy between the annotators were mainly discussed.

The second presenter, Yukio Tono (Tokyo University of Foreign Studies) and Yasutake Ishii (Seijo University), have searched for grammatical items as criterial features of CEFR levels in L2 spoken and written production. We had a discussion regarding missing preposition in learners' written production and grammatical categories which were used as a framework for the study. In addition, the key criteria that should be employed when selecting an appropriate part-of-speech tagger were discussed. It would be ideal to automatically convert erroneous

learner language into corrected target language before inserting the part-of-speech information, but this process cannot be implemented appropriately until a computer-aided error detection and correction system is fully developed. Two of presentations in this session have clearly presented how we can use learner corpus for further refined language acquisition studies.

■講演 (要旨の再掲)

Measures of Productivity and Lexical Diversity

Stefan Evert (Friedrich-Alexander-University of Erlangen-Nürnberg, Germany)

Quantitative measures of productivity and lexical diversity—such as the type-token ratio (TTR), Baayen's productivity index P or Yule's K—play an important role in many corpus studies. They have been used to assess the degree of morphological productivity, to estimate the size of an author's vocabulary, to investigate stylometric differences between writers and settle questions of disputed authorship, to study diachronic changes in grammar, to assess the readability and difficulty level of a text, to explore the linguistic correlates of dementia, and as a feature in the multivariate analysis of linguistic variation.

However, most of the approaches and quantitative analyses found in the literature suffer from serious methodological problems: (i) productivity measures often are sensitive to text size, the presence of lexicalized types and other confounding factors; (ii) there are no well-established methods for assessing the significance of observed differences in productivity, especially in the light of repetition effects due to the non-randomness of natural language; and most importantly, (iii) quantitative measures usually lack a clear linguistic interpretation that links them to intuitive notions of productivity.

In my presentation, I will show how these issues can be analyzed systematically with the help of simulation experiments based on statistical LNRE models. I will also suggest improved approaches and measures that overcome some of the problems and highlight open questions for future research.

■シンポジウム (要旨の再掲)

Issues on multi-word units (MWUs) and collocation

Chair: Laurence Anthony (Waseda University)

On the Creation of a Large-Scale Multi-Word Unit

Resource for Learners of English for Academic Purposes

James Rogers (Meijo University)

Previous research indicates that there are gaps in the literature in regard to a methodology of identifying high-frequency multi-word units (MWUs) for general English purposes, and specifically, English for academic purposes (EAP). Thus, there is also a lack of large-scale resources. In this talk, I present a study in which a novel methodology used to identify high-frequency MWUs of general English is applied to create a similar large-scale resource for EAP. First, the most frequent 500 lemmas in an academic vocabulary list were utilized in the search for lemmatized collocates. Then, these lemmatized collocates were used to identify commonly occurring EAP MWUs, leading to the creation a large-scale EAP MWU list. This results of this study confirmed the importance of native speaker judgments when relying upon corpus data to create a list of MWUs for second language learners that is used to improve their EAP fluency. The results also shed light on the importance of manual checking of corpus data, and the type of low-value items that only manual checking can identify. Most importantly, the study has also resulted in a large-scale EAP MWU resource that not only fills a major gap in the literature, but also confirms previous findings and potentially leads to new discoveries in regard to MWU identification.

Applying a bundle-move connection approach to the development of an online writing support tool for research articles

Atsushi Mizumoto (Kansai University)

Achieving a high level of English proficiency requires a comprehensive English vocabulary of which multi-word units (MWUs) are a critical component. However, acquiring and using these MWUs poses a formidable challenge for second language users of English. In order to facilitate the learning of these units, various online reference resources based on different types of corpora have been developed in recent years. Also, there is a growing interest in resources that are specifically designed to help learners develop an understanding of MWUs above the level of the sentence. In this talk, I introduce some of the current resources available for accessing MWUs that can help to

develop rhetorical competency. Specifically I will focus on a data-driven and theory-based practical writing support tool for research articles (RAs) called AWSuM. This innovative, web-based tool is powered by a combination of rhetorical moves and lexical bundles. It also has an auto-complete feature that suggests the most 11 frequent lexical bundles in a move within an RA section. AWSuM was developed as a proof-of-concept of the bundle-move connection approach. Preliminary user feedback was positive overall, and the writing support tool was found to bring about beneficial effects that genre writing pedagogy explicitly aims to achieve. In light of these findings, the pedagogical implications of the developed tool are discussed, with particular focus on the potential role that it can play in the teaching and learning of technology-enhanced genre writing.

Collocational patterns beyond word pairs

Stefan Evert (Friedrich-Alexander-University of Erlangen-Nürnberg, Germany)

While there is a substantial body of work on the identification and lexicographic description of collocational word pairs as well as idiomatic multiword expressions, only a few studies have addressed longer non-idiomatic word combinations (MWCs). Such MWCs can include collocational patterns involving three or more lexical items that form a series of semantically related MWCs, for example, “set a {dangerous | bad | unfortunate | damaging} precedent”). They also include grammatical constructions with marked lexical or semantic/morphosyntactic preferences, such as the ditransitive use of “earn” (“sth earns sbdy sth”), where the direct object is almost always selected from a narrow semantic field (“nickname, reputation, title, ...”). In this talk, I present ongoing research towards a description of MWC phenomena and the automatic identification of MWC candidates. This approach builds on two premises: 1) Co-occurrence patterns between words cannot be reduced to a one-dimensional association score, but comprise multi-faceted aspects including frequency, salience, and the type-token distribution of each slot; and 2) The complex interrelations between different slots of a MWC can be modelled in terms of nested hypothesis tests, taking into account both significance and association strength. Such nested hypotheses may also involve semantic or morphosyntactic restrictions on the slots, or test whether a larger MWC is composed of overlapping

smaller MWC (e.g. “earn good money” from “earn money” + “good money”).

■ワークショップ1

「言語研究のための Word2Vec 入門」

内田諭 (九州大学)

近年の自然言語処理の技術の発展は目覚ましい。その裏にはコンピュータの飛躍的な進歩と言語情報の大規模な電子化がある。これはコーパスを使った言語研究の実施には恵まれた状況であるといえるだろう。その一方で、多くの技術や技法が溢れるが故に、特に文系の研究者にとって自身の研究テーマに沿ったものを選択することが難しく、日進月歩の進化についていくことは容易ではないという現状がある。

本ワークショップでは、多くの自然言語処理の研究で用いられるようになった word embedding (単語埋め込み) について基礎的な概念を説明し、その中でも特に広く利用されている Word2Vec について概観した。word embedding は単語の意味をベクトル (数字の集合) として扱う技法で、意味の似た単語は似た文脈で用いられるという分布仮説が背景にある。単語の意味をベクトルで表すことで、単語と単語の距離を測定することが可能となり、類義語の抽出やベクトルを用いた意味の演算が可能となる。これらの性質を用いることで品詞解析や機械翻訳、語義曖昧性解消などの精度が向上することが報告されており、自然言語処理において基幹的な技術の一つとなりつつある。Word2Vec は Google の研究チームが開発した word embedding のアプリケーションの一つで、Python や TensorFlow などの広く使われているプラットフォームを通して比較的容易にかつ高速に実行することができるものである。本ワークショップではウェブのインタフェース (Python CGI) を用意し、Word2Vec を実際に動作するところを示した。例えば、約 1 億語のコーパスを用いて作成したモデルを用いて language とコサイン類似度の近い単語を探ると、vocabulary, discourse, culture などがリストされた。また、言語分析を実施する際の注意点として、反意語や多義語などに関する問題点を指摘した。さらに、Word2Vec のモデルを構築する際に最適なコーパスサイズを探索した研究や Word2Vec を使って年代別の文学作品を比較した研究事例を紹介し、さらなる研究の可能性を示した。

■ワークショップ2

「コーパス研究の作法 — 語法研究・辞書編集を通して—」

井上永幸 (広島大学)

本ワークショップでは、特にコーパス研究初心者が遭遇しそうな状況を想定しながら、問題解決の際の糸口を提案していった。まず前半はコーパス基盤的手法の観点から、a man of sense という表現の辞書における扱い方を検討する際の問題点、譲歩を表す if (...) or not を検証する際の注意点、ice pick や grab ... lapel(s) を例に日常生活語彙の用例採取の際に生ずる問題点、インターネットを簡易コーパスとして活用する際に規範的立場からの検証なのかグローバルな英語の立場からの記述的検証なのか目的を明確にしておく必要がある点などを示した。

後半はコーパス駆動的手法の観点から、MI-score について再考した。MI-score は、低頻度のものにも反応するため、その取っ付きにくさから敬遠されがちであるが、逆にキーワードと連想関係の強い語を高頻度のものから低頻度のものに至るまで焦点を当ててくれるので、キーワードの特性を見事にあぶり出してくれる。無論、MI-score が高くても頻度も高い方が信頼度も高まるので、そういった語から優先的にアプローチするのが正攻法である。名詞関連では idea を例に have not [not have] the slightest [faintest, etc.] idea といった定型表現を特定したり、副詞関連では expensive を例に prohibitively [hugely, etc.] expensive といった連想度の高い副詞の特定方法を示した。また、deny との連想度の高い副詞を求める際に、MI-score では strongly, vehemently, etc. といった様態を表す副詞だけでなく、brilliantly, unfairly, etc. といった話し手・書き手の評価を表す副詞を類似の統計値より確認しやすいことを示し、MI-score の機微にふれた。また、最後に he や she と近接生起する MI-score の高い -ly 副詞を紹介し、ジェンダー研究における活用の可能性にも言及した。

<第 45 回大会発表者募集>

英語コーパス学会第 45 回大会は、2019 年 10 月 5 日 (土) と 10 月 6 日 (日) に高知県立大学で開催されます。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, 質疑応答10分)

【発表言語】英語または日本語とします。英語での発表を推奨します。

【応募方法】発表申込ウェブフォーム(<https://goo.gl/forms/ObUe2d08Ek1FrXQp2>)に必要事項を記入・送信の上, 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。先行研究で述べられていることと, 自身が明らかにしたこと・自身の意見を明確に区別してください。

使用したコーパスやデータを明確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。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ります。

英語で発表する場合は, 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。それ以外の注意事項・書式指定は日本語での発表の場合と同じです。

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。

【応募期限】2019年5月31日(金)必着

【採否決定】2019年7月初旬(予定)

<理事会の決定事項について>

10月5日(金)17時30分より東京理科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報告いたします。

(1) 事務局

・会計(再任)

宇佐美裕子先生(東海大学)

(2) 編集委員会

・委員(新任)

今林修先生(広島大学)

小島ますみ先生(岐阜市立女子短期大学)

<会誌『英語コーパス研究』第27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
田畑智司(大阪大学)

『英語コーパス研究』第27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】2019年11月30日(土)

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】2020年1月

【発行日】2020年3月31日

(発送は2020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>

学会賞選考委員会 委員長
西村秀夫(三重大学)

2019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、2019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。

※インターネット上で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】2019年6月30日（日）

【審査結果の報告および表彰式】第45回大会総会（高知県立大学 10月5日）

<2019年度春季研究会について>

2019年4月20日（土）に名古屋工業大学にて2019年度春季研究会を開催いたします。現在ある5つの研究会（SIG）が、30分ずつ各自の内容で研究発表・SIG紹介等を行います。今回は5つの研究会が連続して発表等を行いますので、全ての研究会の話をお聞きいただけます。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。

す。

<新入会員紹介>

竹中裕貴	（島根大学）
相吉晃太郎	（東京大学, S）
桐原浩昭	（比叡山高校）
武内梓朗	（九州産業大学）
Joe Geluso	（Iowa State University, S）
川本渚凡	（東京外国語大学, S）
萩原明子	（東京薬科大学）
西出公之	（奈良大学）

（Sは学生会員）

（2018年7月6日から2018年12月16日の入会者）

<事務局からのお知らせ>

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

2018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、2018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

■The 4th Asia Pacific Corpus Linguistics Conference (APCLC 2018)に参加して
杉森直樹 (立命館大学)

2018年9月17日から19日まで香川県の高松市で開催された APCLC 2018 に参加した。APCLC は、APCLA (Asia Pacific Corpus Linguistics Association)が2年毎に開催している国際会議であり、第1回大会が2012年に New Zealand の University of Auckland で開催された後、第2回が2014年に Hong Kong Polytechnic University で、第3回が2016年に北京の Beihang University (北京航空航天大学) で開催されており、今回はそれに続く第4回の開催であった。大会自体は3日間であったが、前日には Pre-conference Workshop が開催され、大会終了後には Excursion も実施されたため、延べ5日間の日程となった。台風21号の被害で関西国際空港への国際便に影響が出ていたことが心配されたが、アジア諸国を中心に海外からも多くの参加者があった。会場となったサンポート高松の国際会議場はガラス張りの近代的な建物で、高松港に面した会場からは瀬戸内海の美しい眺めを見ることができた。

大会前日の16日には以下の4件の Pre-conference workshop が開催された。Workshop 1: Multivariate analysis with R (Stefan Evert), Workshop 2: WordSkew: Text and Corpus (Michael Barlow), Workshop 3: Exploring Features of WordSmith 7.0 (Mike Scott), Workshop 4: The Prime Machine – a tool to help English language learners get started with the exploration of English language and literature (Stephen Jeaco)。私は Workshop 1 と 3 に参加したが、どちらも会場のコンピュータ教室を使用して hands-on 形式で行われた。Workshop 1 は R の使用方法だけでなく、多変量解析を行う際に注意すべき点についての解説も行われ、大変勉強になった。また、Workshop 3 では WordSmith の最新バージョンを使用してコーパスデータの分析を体験することができ、WordSmith の有用性、多機能性を改めて実感することができた。

Plenary speech としては以下の5件の講演があった。Plenary speech 1: “Corpus linguistics and the study of fiction – methodological and theoretical challenges” Michaela Mahlberg (University of Birmingham, UK), Plenary speech 2:

“Understanding and Advancing the Data-Driven Learning (DDL) Approach” Laurence Anthony (Waseda University, Japan), Plenary speech 3: “Distributional Methods in Corpus Linguistics: Towards a Hermeneutic Cyborg” Stefan Evert (Friedrich-Alexander-Universität, Germany), Plenary speech 4: “How Deep Learning Changes Natural Language Processing” Naoaki Okazaki (Tokyo Institute of Technology, Japan), Plenary speech 5: “The Individual and the Group in Corpus Linguistics” Michael Barlow (University of Auckland, New Zealand)。コーパス文体論から自然言語処理まで幅広いテーマの講演が行われ、本学会会員の Laurence Anthony 先生も Data-Driven Learning (DDL) のテーマで講演をされたが、どの講師の方も豊富な PowerPoint スライドを用いて話を進められ、専門外のテーマの話であっても理解がしやすくなるような工夫がなされていた。

研究発表は107件の発表があり、コーパス言語学に関する様々な研究テーマの発表が行われたが、学習者コーパス、DDL、EAP、談話分析に関する発表が比較的多かった印象である。英語コーパス学会も本大会の後援団体となっていたこともあり、会員からは以下の32件の発表があった(プログラム順に記載)。“Acquisition of Tense/Aspect markers in Learners Corpora of English/Chinese/Japanese” (Keiko Mochizuki, Yaming Shen, Zhang Zheng and Laurence Newbery-Payton), “The effects of corpus use on correction of article- and preposition-omission errors” (Yoshiho Satake), “Errors and Beyond—A Corpus-based Stylistic Analysis of “Japanese English” Discourse” (Emi Izumi), “Is the Continuation of The Mystery of Edwin Drood a Posthumous Work of Charles Dickens?: A Multivariate Analysis” (Katsumi Goto), “A corpus-driven analysis of present-tense fiction” (Reiko Ikeo), “Visualizing English Classroom Spoken Data on Multi-modal Interface to Create Versatile Linguistic Resources” (Noriaki Katagiri and Yukiko Ohashi), “Predicting EFL learners’ oral proficiency levels in monologue tasks” (Yuichiro Kobayashi, Yusuke Kondo and Mariko Abe), “Functional Classification of Lexical Bundles in TED Talks” (Naoki Sugimori), “Exploring word-formation in science fiction using a small corpus” (Dax Thomas), “How distinctive is your corpus? A metric for the degree of deviation from the norm” (Yo Sato and Kevin Heffernan), “Lexical characterization of semi-popularization articles on agricultural topics” (Kayo Yamamoto, Tamao Araki and Richard S. Lavin), “Inter-annotator agreement:

By hook or by crook” (John Blake), “Developing multilingual language learning resources using the CEFR-J” (Yukio Tono), “Distinguishing L1 and L2 Using Three Linguistic Aspects: A Logistic Regression Model Study” (Masatoshi Sugiura, Yoshito Nishimura and Daisuke Abe), “S-genitives and Of-genitives Seen in L2 English Learners’ Essays: A Study Based on the ICNALE Written Essays” (Shin’Ichiro Ishikawa), “A Japanese/English DDL Tool for Primary School CEFR Pre-A1 EFL Learners” (Chikako Nishigaki, Kiyomi Chujo and Shiro Akasegawa), “A corpus-based analysis of the use of subordinators by Japanese learners of English” (Hikaru Misumi and Yukio Tono), “Annotating the functions of learner utterances from a spoken corpus and assigning the degrees of their grammatical accuracy and discoursal acceptability” (Aika Miura), “Corporate Cautionary Statements (Disclaimers): A Critical Genre Analysis of Professional Communication” (Chie Urawa), “CEFR Receptive and Productive Vocabulary Knowledge of Japanese English Learners” (Hiroko Usami), “Investigating Japanese EFL learners’ overuse/underuse of English grammar categories and their relevance to CEFR levels” (Yasutake Ishii and Yukio Tono), “Assigning CEFR-J levels to English texts based on textual features” (Satoru Uchida and Masashi Negishi), “A Corpus-based Study on Frequent Noun Phrases in Engineering Academic Texts” (Yuka Ishikawa), “The Use of Active Reading to Integrate Language Skills: A Data-Driven Learning Approach” (Kunihiko Miura), “Prefixes and suffixes in Japanese junior high school English textbooks” (Mitsuhiro Morita, Satoru Uchida and Yuka Takahashi), “A Usage-based Approach to Criterial Canonical Construction in an L2 Learner Corpus” (Yoshiyuki Notohara), “Examining the applicability of the mean dependency distance (MDD) for SLA: A case study of Chinese learners of Japanese as a second language” (Saeko Komori, Masatoshi Sugiura and Wenping Li), “A phraseological account of structural patterns: ‘it is no NN that’ and ‘there is no NN that’” (Atsuko Umesaki), “An Analysis of the Vocabulary Used in Oxford Reading Tree: As a Reference for Early English Education in Japan” (Shuji Hasegawa), “Extracting Patterns from Transition of Occurrence Frequency of Grammar Items in a Junior High School Textbook in Japan” (Amma Kazuo), “The varying complexity of the syntactic role of nouns: Evidence from Japanese corpora” (Kevin Heffernan and Yo Sato), “Historical Analyses of English

Negative prefixed words from ME to PE - ‘-able’ adjectives-“ (Akira Okada), “Negation in Benjamin Franklin’s Writings: A Stylistic Analysis of his Autobiography and Letters” (Yoko Iyeiri).

懇親会は高松市の代表的な観光名所である栗林公園内の商工奨励館で開催されたが、栗林公園の散策ツアーも行われ、公園内の見事な庭園風景を堪能することができた。懇親会前には二胡奏者の WeiWei Wu さんの演奏会が開催され、徐々に日が暮れてゆく中で幻想的なひとときを過ごすことができた。大会終了後の Excursion では、瀬戸大橋主塔の塔頂への登頂、讃岐うどんの手打ち体験、琴電の車両貸し切りランチなどが開催され、好評であったようである。

今回の APCLC 2018 には高松観光コンベンションビューローの後援があったとのことで、会場の設備や懇親会が大変充実していたのが印象的であった。本学会からも多くの参加者があり、このような規模のコーパス言語学の国際大会が日本で開催されたのは大変意義深いことだと思われる。Conference Chair を務められた投野会長をはじめ、大会に協力された皆様にはお礼を申し上げたい。なお、次回の大会は 2020 年 2 月 10～13 日に韓国ソウルの Yonsei University (延世大学) で開催予定とのことであった。

*APCLC 2018 については <https://apclc2018.org/> を参照されたい。

◇Forum の原稿を随時募集しています

英語コーパス学会 Newsletter では会員の皆様からの Forum への投稿を募集しています。国際学会報告、研究会の紹介、新刊紹介など、会員の皆様の情報交換の場として Forum が活用されることを願っております。以下、詳細を記します。

- Forum のテーマ：国際学会報告、研究会の紹介、新刊紹介など英語コーパス学会にとって有益と思われる情報
- 締め切り：5 月末あるいは 10 月末
- 分量：800–1,600 字程度（写真等の画像の掲載も可能です）
- 送付先：jaecs.hq@gmail.com

2018年12月29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/>
