

英語コーパス学会 Newsletter No. 87

Dec. 29, 2019

■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

NL87号の概要

- [1] 第45回大会報告
- [2] 2020年度春季研究会・総会予告
- [3] 第46回大会予告・発表募集
- [4] SIGの催し
- [5] 理事会報告
- [6] 事務局報告
- [7] 会誌投稿募集
- [8] 学会賞・奨励賞推薦募集
- [9] FORUM
- [10] 第45回大会講演・シンポジウム・発表・ワークショップ概要

<1. 第45回大会報告>

■概要

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)の2日間にわたり高知県立大学(永国寺キャンパス)にて開催されました。第45回大会では、Tony McEnergy氏(Lancaster University)の講演、井佐原均氏(豊橋技術科学大学)によるワークショップ、投野由紀夫氏(東京外国語大学)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。20年振りの四国での大会ということもあり、発表数も多く、活発な大会となりました。また、シンポジウムと講演に加え、研究発表の2つのセッションが英語で行われるなど、学会の国際化が着実に進んでいることが認識できる大会でもありました。

1日目午後の開会式では、まず投野由紀夫会長による開会の挨拶に続き、開催校である高知県立大学の副学長であり英語コーパス学会の会員でもある五百蔵高浩氏にご挨拶をいただきました。

次に、石井康毅事務局長(成城大学)の司会で総会が行われました。はじめに、役員人事についての報告がなされました。(詳細は後述の「<5. 理事会報告>」のセクションをご覧ください。)続いて、功労会員制度の新設、副会長と会計の人数変更、そして若干の文言の変更か

ら成る会則改定案が示され、承認されました。また、会計の宇佐美裕子氏(東海大学)より、2018年度会計報告及び2019年度予算案が示され、いずれも承認されました。総会の最後に、投野由紀夫会長から学生優秀発表賞規程の新設と奨励賞の対象者変更について報告が行われました。

総会に引き続き、学会賞選考委員長の西村秀夫氏(三重大学)より学会賞審査報告が行われ、本年度の奨励賞が内田諭氏(九州大学)に授与されることが発表されました。この授賞は「リーディング・リスニングテキストのCEFRレベル判定ツールCVLAの開発と公開」に関する内田氏の一連の研究に対するものです。なお、本年度の学会賞の該当者はありませんでした。

奨励賞の授賞式(左が内田氏、右が投野会長)



午後の研究発表では4つのセッションが置かれました(2室で並行実施)。第1セッションは仁科恭徳氏(神戸学院大学)、第2セッションは菅原崇氏(岐阜工業高等専門学校)、第3セッションは杉森直樹氏(立命館大学)、第4セッションは内田諭氏の司会のもと、計10件の研究発表が行われました。その後、Tony McEnergy氏による講演「Revisiting A Corpus—the BNC 2014」が行われました。

2日目の午前には2つの研究発表セッションが置かれ、第5セッションは木山直毅氏(北九州市立大学)、第6セッションは能登原祥之氏

(同志社大学)の司会のもと、計4件の研究発表が行われました。

発表に引き続き、井佐原均氏によるワークショップ「人工知能とデータ-自然言語処理とコーパスを例に」が行われました。

2日目の午後は、Yukio Tono氏の司会進行で、Yukio Tono氏、Shin'ichiro Ishikawa氏、Hitoshi Isahara氏、Tony McEnery氏によるシンポジウム「Gazing into a crystal ball: what you can see in the future of corpus linguistics」が行われ、コーパス言語学の未来についてのディスカッションが行われました。発表後には質疑応答の時間も設けられました。

第45回大会は2日間を通して、85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。

本NLの末尾には、講演、シンポジウム、ワークショップの概要(大会資料からの再掲)、各発表者が執筆した研究発表の概要、そして司会者による各セッションのまとめを載せております。

<2. 2020年度春季研究会・総会予告>

下記の通り、2020年度の会員総会および研究会を行います。多数のご参加をお待ちしています。

英語コーパス学会 2020年度春季研究会・総会

日時：2020年4月18日(土) 13:30~17:10
会場：関西大学(千里山キャンパス岩崎記念館)(阪急関大前駅下車徒歩5分程度)
参加費：会員無料、非会員1,000円
参加申込：不要。直接会場へお越しください。

プログラム(変更の可能性あります)

13:15~ 受付

13:30~13:50 会長挨拶・会員総会

14:00~15:30 ツールと統計手法研究会特別シンポジウム(90分)

Chair: Laurence Anthony (Waseda University)

Presentation 1

Revisiting word counting units in the development of word lists

Atsushi Mizumoto (Kansai University)

Presentation 2

To Be Announced

15:40~17:10 語彙研究会特別シンポジウム(90分)

内容は後日お知らせします。

17:30~ 懇親会

(※同日の10:30~13:00に理事会/役員会を開催します)

<3. 第46回大会予告・発表募集>

第46回大会は、以下の通り開催予定です。ぜひ早めにご予定ください。

英語コーパス学会第46回大会

日時：2020年10月3日(土)・4日(日)

会場：早稲田大学(西早稲田キャンパス)

(※大隈講堂のある早稲田キャンパスではありませんのでご注意ください)

講演他については、アントニ・ロレンス大会実行委員長の下、現在、企画中です。決まり次第、ウェブサイトやニュースレターでご案内します。

また、上記大会での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)

【英語発表の推奨】学会の国際化推進のため、英語による発表を推奨しています。

【応募方法】期限までに以下の2つの作業を完了する必要があります。

(A) 発表申込ウェブフォームに必要事項を記入・送信する。

(B) 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)を利用して発表概要をWord形式で作成し、電子メールの添付ファイルで送付する。メール本文には代表者名と発表題目を明記する。

ウェブフォームと概要の送付先は、決まり次第学会ウェブサイトに掲載します。

【発表概要作成要領】

・冒頭には題目のみ記し、発表者名は書かない。

・概要本体（題目・文献リストを除く）は、日本語の場合、800～1,200字、英語の場合は400～600語で作成する。

・必要に応じて参考文献を明示してよいが、文献リストの部分は字数に含めない。

・ツールやコーパス開発などの発表を除き、リサーチクエスチョンを示した上で、研究から得られた知見を明快に記述する。

・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する。

・使用したコーパスやデータを明確に記述する。

・公平な審査を行うため、本文・文献とも、応募者の特定につながる情報を記載しない。

・脚注は使用しない。

【審査】審査は応募者名を伏せたまま、提出された発表概要のみに基づいて行います。提出された発表概要に応募者の特定につながる記述が認められた場合、編集して審査に付す場合があります。

【要旨・報告の作成】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】2020年5月31日（日）必着

【採否決定】2020年7月初旬（予定）

<4. SIGの催し（2020年1～3月）>

英語コーパス学会のSIGにより、下記の研究会が開催されます。SIGメンバーに限らず、本学会会員はどなたでも参加いただけます。

ESP研究会 2019年度研究会

「1日でCasualConcを極める新春ワークショップ」

日時：2020年1月5日（日）9:30～16:30

会場：京都キャンパスプラザ（JR京都駅前徒歩3分）

講師：今尾康裕氏（大阪大学）

参加費：無料

参加申込：<https://t.ly/7pWYyE>（会場の都合上、事前申し込みをお願いします）

問合せ：ishikawa.yuka@nitech.ac.jp 石川有香

語彙研究会 2019年度研究会

日時：2020年1月25日（土）14:00～17:00

会場：中央大学 後樂園キャンパス（6号館6701教室）

詳細：<https://jaecs-lexsig.blogspot.com>

参加費：無料

参加申込：不要。直接会場へお越しください。

問合せ：sugimori@mbox.kyoto-inet.or.jp 杉森直樹

プログラム

発表1：14:00～14:30

杉森直樹（立命館大学）

「TED Talksの語彙頻度プロファイル：JACET8000を用いた分析」

発表2：14:35～15:05

石川慎一郎（神戸大学）

「学習者作文の誤り検出：Grammarly(R)と人間の校閲者を比較して」

発表3：15:10～15:40

小屋多恵子（法政大学）

「理系大学院生の英語論文に特有の表現」

発表4：15:45～16:15

日臺滋之（玉川大学）

「言いたかったけれど言えなかった表現コーパス：中高生の発信のために必要となる語彙」

発表5：16:20～16:50

阿部真理子（中央大学）

「中高6年間を通して増える誤り・減る誤り」

17:15～懇親会

<5. 理事会報告>

2019年10月4日（金）17時30分より高知県立大学において理事会が開催されました（一部議案は後日のメール審議で検討・承認）。承認された事項についてご報告いたします。

(1) 人事について

・副会長（～2020年3月）および次期会長（2020年4月～）として石川慎一郎氏（神戸大学）が選任されました。

・理事として内田諭氏（九州大学）が選任されました（2019年12月～）。

<副会長就任挨拶>

石川慎一郎（神戸大学）

このたび、井上永幸先生の後任として、副会長に選任された石川です。副会長としての任期（2020年3月まで）においては、投野会長を支え、各種の仕事を遅滞なく進めてまいりたいと存じます。また、4月以降については、投野会長が手掛けられた改革を引き継ぎ、学会活性化のための施策を立案・実行していきたいと考えております。力不足ではありますが、長く御世話になってきた学会のため、微力ながら最善を尽くす覚悟でございますので、会員の皆様におかれましては、ご指導ご鞭撻のほどどうぞよろしくお願い申し上げます。

(2) 学生優秀発表賞の新設について

英語コーパス学会第46回大会より、学生会員の研究奨励と、学会の活性化を目的とし、専任職を持たない学生会員の大会での単独口頭発表の中から若干名の優秀発表賞を選出することを決めました。

(3) 研究会（SIG）の新設について

6つ目の研究会として、「コーパスによる言語変異研究会」が新設されました（代表：谷明信氏（兵庫教育大学））。現在、10名の会員がおり、2019年度から活動を行う予定です。

<6. 事務局報告>

(1) 会費納入のお願い

2019年度および過年度の会費（一般6,000円、学生3,000円）について、未納がある方は、至急、納入をお願いします。一定期間の未納があると、会員資格が停止されます。

2020年度会費（同上）については、学会からの郵送物に合わせて請求書を同封する予定ではございますが、それをお待ちいただく必要はございません。2020年4月以降、できるだけお早めにお納めいただけますと幸いです。

お手元に会費払込取扱票がない場合は、郵便局に備え付けの払込取扱票に以下を記入の上、ご送金ください。

●振替口座：00930-3-195373

●加入者名：英語コーパス学会

※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こととなります。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。

(2) 新年度の所属・住所等の変更

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。

(3) 新入会員紹介

藤田郁（立命館大学, S）

河本健（広島大学）

（Sは学生会員）

（2019年7月19日から12月7日までの入会者）

<7. 会誌投稿募集>

『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28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】2020年11月30日（月）

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】2021年1月

【発行日】2021年3月31日

（発送は2021年5月下旬の予定）

<8. 学会賞・奨励賞推薦募集>

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）

2020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、2020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、大学院修士課程または博士前期課程修了後15年未満の者で、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】2020年6月30日（火）

【審査結果の報告および表彰式】第46回大会総会（早稲田大学（西早稲田キャンパス））

<9. FORUM>

(1) 新刊紹介

西村秀夫（編）（2019）『コーパスと英語史』（英語コーパス研究シリーズ第6巻）ひつじ書房

報告：柴崎 礼士郎（明治大学）

2015年10月1日から刊行が開始された「英語コーパス研究シリーズ」が、2019年10月2

日刊行の第1巻をもって完結した。第1巻には、コーパス言語学の研究概観（第1章）、英語コーパス学会初代会長齊藤俊雄氏（大阪大学名誉教授）の遺稿（第2章）、更に、多様な研究背景を持つ15名による「私のコーパス利用」および総索引が含まれている。『日本英文学会九州支部第27回大会』（於：熊本県立大学、2019年10月26-27日）では、本シリーズ全巻刊行成果を踏まえて、新たな英語コーパス研究の可能性が示されている（シンポジウム「これからの英語コーパス研究」）。

全巻の導入とも言える第1巻を除くと、刊行までに最も時間を要したのが第6巻『コーパスと英語史』であるが、他巻に劣らぬ内容に仕上がっている。第1章（編者・西村秀夫、以下敬称略）では、電子化以前のものを含む代表的なコーパスやコンコーダンス、および、国内外の史的英語研究が鳥瞰できる内容である。以下、字数制限により、BNCなどのスペルアウトは極力省いてある。

第2章（塚本聡）での考察は、PPCME2を用いた中英語における二重目的語構文の語順精査であり、間接目的語（与格）に付与されていた内在格の衰退過程にも触れている。主節、従属節および不定詞節に構造を分け、それぞれに生起する動詞と二種類の目的語の語順を調査した結果、どの時代でもV IO DOが優勢であることが分かった。人称代名詞が目的語の場合には別傾向が確認できる点、PPCME2の解析情報ミスを指摘するほどの事例精査などは説得力がある。構文解析情報付きコーパスを利用することで、先行研究の更新を可能とした点など学ぶべきところが多い。

第3章（谷明信）では、ICAMETを用いて、認識動詞witの動詞性消失の過程と談話機能標識化を中英語全体で考察している。命令形としての生起率が高いことや、競合するknowに優位に立たれる推移を地域毎に明示するなど、本書の中で最も専門性が高いと感ぜられた。200以上もある異綴形をどのように検索、抽出、分析するかについても触れられている。「刊行のことば」を借りれば、歴史的に「英語研究に携わっている方々」には特に示唆に富む論考であろう。

第4章（家入葉子）では、サンプリングと代表性を意識しつつ著者が開発したEMEPSを用いて、否定を含意する動詞（forbid, doubt, avoid, prohibit）の補部選択の史的変遷を調査している。全体として「that節>to不定詞>動名詞」という変化の方向性が確認でき、著者自身がOEDを用いて行った別の調査結果とも概ね一

致している。考察結果と同様に重要なのは、EMEPS の利用価値が損なわれていない反面、該当事例数が必ずしも十分でなかったことなど、コーパス編纂がはらむ問題点を著者が咀嚼反芻している冷静さである。

第5章(大津智彦)では、1800年代のイギリス英語、1990年代のイギリス/1990年以降のアメリカ口語英語における現在完了形の浸透度を考察している。前者には The Project Gutenberg からの15文学作品を、後者には BNC の spoken component と The Screenplay Database からの映画演劇に加え、CNN の Larry King Live (2009年1月放送分)を用いている。構文毎に過去形から現在完了形への移行に違いが見られることや、アメリカ英語では過去形が選択される傾向が確認できた点など、英語変種毎の推移の違いが提示されている。

第6章(水野和穂)は、後期近代英語研究の現状紹介にはじまり、ヘルシンキ大学が提供する CoRD (Corpus Resource Database)の中から代表的な後期近代英語に関するコーパスを詳述しており、学習者目線の出だしと言える。その後、PPCMBE と ARCHER を用いて、異綴り字 (e.g. *show/shew*), 語法 (e.g. *you was*), 両コーパスの代表性の問題を具体的に提起し、更に、構文解析情報付きである PPCMBE の有効利用の案として、「特殊目的コーパス」の編纂にも触れている(第4章の EMEPS, 第1章で指摘されている PPCME2 の解析情報ミスとも関係する)。

第7章(内田充美)は、BYU コーパス群の詳細な解説と利用により、「嫌われ表現」(e.g. *at the end of the day, the bottom line is*)の発達推移に取り組んでいる。COHA を用いた長期的考察も興味深い。COCAなどを駆使した直近の変化とジャンル間の比較考察が目を開く。話しことばでの使用が全体的な傾向ではあるが、*proactive* などの拡大傾向は書きことばが中心のようである。検索条件の調整などの示唆にも富んでいる。

各章を通じての惜しめない情報提供は本書の最大の魅力である。コーパスを駆使した場合でも、最終的には手作業による分析が重視されている点にも深く共感できる。古英語に正面から取り組む論文や、音声データで辿る英語史なども加わると更に興味深かったかもしれないが、別稿として期待したい。脱稿から刊行までに6年あまりを要した点は研究の「鮮度」面で残念であるが、初学者にも分かりやすい構成と、誤字脱字も極めて少ない内容は、予想以上の波及効果をもたらすものと信じて止まない。願わく

は、人口減と目される歴史言語研究に歯止めが掛かって欲しい。

最後に、利用可能なコーパスの拡大と普及が、原典テキストの軽視に繋がらぬよう意識したい。言語学者に限らず資料(史料)を重んじる研究者は多く、幕末史の権威である故佐々木克氏(京都大学名誉教授)もその一人である。「史料を読みなおそう(中略)幕末期史料は漢文脈の文章だから漢和辞典は必携である。愚直に辞典のページを繰る日が続いた」と述べている(『幕末史』pp.341-342, ちくま新書, 2014)。人間と AI による共同作業が最高の成果を生み出すという自然科学分野の報告もある。言語研究においても、原典テキストと電子コーパスが共存共栄できる近未来を切に願っている。

(2) 学会報告

CL 2019 (10th International Corpus Linguistics Conference)に参加して

報告: 川本 渚凡(東京外国語大学大学院生)

Corpus Linguistics (通称 CL) は 2001 年にランカスター大学で始まった国際会議である。これまでに、ランカスター大学、リバプール大学、バーミンガム大学がホストとなり、隔年で開催され本大会で 10 回を迎えた。本学会はヨーロッパのみならず、世界各国から多数のコーパス言語学の研究者が一堂に会する。今夏は欧州での記録的な猛暑が日本でも連日ニュースになったが、本学会期間中はヒースロー空港周辺で 36.9 度を記録し、開催地であったカーディフは第一線で活躍する研究者同士の活発な活動が行われ文字通り「あつい」学会となった。

CL2019 は 7 月 23 日から 26 日の日程で、英国ウェールズにあるカーディフ大学 (Cardiff University) で開催された。基調講演としては以下の 5 件があった: Opening Plenary: Corpus Linguistics in the Digital Humanities (Svenja Adolphs, the University of Nottingham), Plenary Speech 1: The present and future of Corpus Linguistics and health(care) communication (Elena Semino, Lancaster University), Plenary Speech 2: Forwards, backwards, sideways? Directions of interaction between CL and EAP (Hilary Nesi, Coventry University), Plenary Speech 3: A Question of Time (Tony McEnery, Lancaster University), Plenary Speech 4: The use of corpus linguistics in analyzing scripted fictional media (Bróna Murphy, the University of Edinburgh)。

4 日間で 214 件の研究発表、6 件のパネル・ディスカッションが行われた。ポスター発表の

正確な件数はわからないが、学会中 4 回にわたり、"Language Learning and Teaching"と"Corpus methodologies"の 2 テーマに関するポスターが掲示された。発表件数もさることながら、研究テーマも多岐にわたり、"Corpus Methodologies," "Applications of Corpus Linguistics," "Discourse, History and Society," "Language Learning and Teaching"の 4 ストリームと下位分野を広く含むもので、また英語以外の言語の発表も数多くあり、本国際大会がまさしくコーパス言語学の最新動向の showcase という印象であった。

また、筆者は残念ながら参加しなかったが、学会に先立ち、7 月 22 日には Pre-conference workshop として、全日ワークショップが 2 件、半日ワークショップが 8 件、計 10 件という充実ぶりであった。英語コーパス学会会員の発表は、5 件あり、早稲田大学 Laurence Anthony 氏と前年度 JAECS の plenary speaker として招聘された Stephan Evert 氏 (Erlangen University) が共同で "Embracing the Concept of Data Interoperability in Corpus Tools Development" と題して発表した他、Let's look at corpus data with CasualConc! (Yasuhiro Imao), Automatic coding of a team leader's eye-gaze in emergency care simulation: developing a multimodal corpus with the sceneAnalysis GUI (Keiko Tsuchiya, Takeshi Saitoh, Kyota Nakamura, Takuma Sakai, Takeru Abe and Akira Taneichi), Epistemic modal verbs and adverbs in Japanese university students' academic writing (Kazuko Fujimoto) などの会員諸氏の発表があった。

また、筆者自身も主任指導教授である投野由紀夫先生 (本学会会長) と共同研究で "Predicting CEFR levels of illustrative examples: a corpus approach" の発表の機会を得た。まだまだ駆け出しの研究者であるが、これまで国内外いくつかの学会に参加し、発表する機会を得た。その中でも、audience の質問やコメントにとっても温かさを感じる学会であった。まだまだ予備的な段階にある研究で、結果も期待通りではなかったのだが、聴衆から複数の建設的なコメントをいただき大変感謝している。発表後にも使用するデータや方法論についての助言や共同研究のお誘いをいただいた。大規模な学会で最大 8 トラックの平行・セッションで、興味のある発表全てを聞けないのが何とも残念であったが、英語コーパス学会の会員のみならず全員が興味を持たれる発表がきっと見つかる学会だと思う。なお、次の大会は 2021 年にアイルランド リムリックで開催される。

※ CL2019 については、<http://www.cl2019.org/> をご参照ください。

◇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

<10. 第 45 回大会 講演・シンポジウム・発表・ワークショップ概要>

第 1 日第 1 セッション

[司会] 仁科 恭徳 (神戸学院大学)

Sentential adverbs as connectors between moves in the results sections of basic medical research articles

Tatsuya Ishii (Hiroshima University,
Postgraduate Student)

Takeshi Kawamoto (Hiroshima University)

Members of a discourse community share moves, units of meaning that writers have learned to use in an organized way. The organization of moves in a specific environment is described by move analysis (Swales 1990; Dudley-Evans & John 1998). The analysis reveals how the development of meaning in a given community is realized by phraseological patterns as lexical units (Hyland 2009; Hunston, 2013). The move descriptions produce useful corpus data to uncover the link between the conventionalized phraseology characteristics of specialist genres and the function of moves (Biber et al. 2007). The findings will be relevant especially to the analysis of medical research articles (Nwogu 1997; Gledhill 2000; Kanoksilpatham 2005; Saber 2012). To date, the previous studies for identifying the typical phraseologies have observed the concordance lines of the prepositions in keywords, as in Gledhill (2000) or the nouns such as shell nouns as in Schimid (2000), or have illustrated 4-gram sequences, consistent with Saber (2012). However, little is known about the characteristics of the behavior of

adverbs as keywords and the function of sentential adverbs as connectors between moves. Phraseologies as connectors between moves play a significant role in conveying the communicative purposes of writers. This phraseological study demonstrates that sentential adverbs function as connectors between moves in the results sections of basic medical research articles. The analysis of sentential adverbs reveals that phraseologies associate strongly not only with the function of moves but also with rhetorical patterns. In collecting corpus data for this study, we collected 304 basic medical research articles in 30 journals published in 2014 (in total approximately 1.6 million words). In addition, we divided each article into 12 moves, where the results section was categorized into three moves in terms of the revised version of Kanoksilpatham (2005): (R1) Restating methodology, (R2) Announcing results, and (R3) Commenting results. Using AntConc, we investigated the concordance lines of 4, 12, and 8 sentential adverbs in (R1), (R2), and (R3), respectively, as move-specific keywords. For example, followed by the phrase *we found that* or *we observed that*, the adverb *Interestingly* or *Notably* in initial sentences as an attitude marker is used as connector between (R1) and (R2). Moreover, preceding the phrase *our data indicates that* or *these studies suggest that*, the adverb *Together*, *Taken together*, or *Collectively* as a summary marker is utilized to connect (R2) and (R3). In our presentation, more phraseologies will be exhibited. Collectively, systematic methodologies focusing on the behavior of sentential adverbs are powerful tools for describing phraseologies strongly

Move Analysis of Hotel Overviews on Official Websites of Hotels in Japan: Luxury Strategies in Overview Text

Yukie Kondo (Ritsumeikan University)

Introduction

Official hotel websites usually place overviews of their hotels on the top page. They are short but have a prominent role considering that an official website is an important means for communication between the hotel and the (potential) guests. Kondo (2018) developed three moves and three steps in overviews of London hotels using the genre analysis framework (Bhatia, 1993; Swales, 1990) and investigated the typical move structure of hotel overviews. These moves were Move 1. Defining

self, Move 2. Establishing features, and Move 3. Establishing connections. Move 2 were further divided into three steps: Step 1. History/architecture, Step 2. Location, and Step 3. Facilities. By comparing the move implementation rates and keywords between the higher-grade and the lower-grade hotels, how the higher-grade hotels used luxury strategies in overviews was discussed. However, whether the results obtained were exclusive to hotels in London or whether they could be applied to hotels in general is still unknown. Therefore, this study takes Japan for another area to examine and analyses overviews of hotels located in Japan. The analyses are conducted by verifying five hypotheses formulated based on the results obtained from the analyses of London hotel overviews.

Corpus

The corpus for this study is composed of hotel overviews on websites of luxury hotels in Japan, which were chosen as 2018 Forbes Star Award Winners of 5-star, 4-star, and Recommended (29 hotels in Japan). Corpus of overviews of London 5-star hotels chosen by The AA Hotel Guide 2016 (47 hotels) comprised by Kondo (2018) was also used to compare the results.

Hypotheses

Summary of the five hypotheses are as follows: Hotel overviews on the websites of luxury hotels in Japan: 1. Have a similar organization to those found in the overviews of hotels in London. 2. Use the luxury strategy of ‘evoking exclusivity’ by not directly addressing or inviting readers. 3. Use the luxury strategy of ‘creating abstractness’ by using abstract expressions in the core of the sentence and state concrete information in modifying phrases or clauses. 4. Use abstract expressions that include ‘dream value’ by using Move 2: Step 1: History/architecture. 5. Have a luxury strategy unique to hotels in Japan.

Results and Discussion

The overviews of luxury hotels in Japan consisted of almost the same organization in terms of the move implementation, but there was a difference in how the hotels in Japan and those in London executed luxury strategies. The difference was the reflection of their culture, especially the different way of showing hospitality. This reflection can be further applied as a strategy using a different way of evoking the readers’ dream. The author

proposes ‘evoking a dream’ as a luxury strategy in texts. For hotels in London, ‘hotel-has-what,’ the hotel’s facility, convenient location, or the hotel’s architecture, and how high in quality they are were the main focus, while for hotels in Japan, ‘hotel/guests-do-what,’ what the hotel offers to the guests, what their association is with the country’s history and culture, and how the hotel’s identity was created in relation to the history and culture were more distinctively described. Even if the organization of the overview is similar, what could be a ‘dream’ can be different.

A Fast, Scalable, Portable Corpus Database Architecture for Small- and Large-Scale Corpus Research

Laurence Anthony (Waseda University)

In recent years, corpus methods have been applied in an increasingly wide variety of disciplines including not only traditional branches of linguistics, but also healthcare, law, environmental studies, finance, and literature studies. The expanding uses of corpus methods have led to an increased need for researchers to access small, carefully-constructed, and heavily annotated datasets as well as very large somewhat opportunistically created datasets, such as those collected from the Web. Online corpus portals such as SketchEngine (www.sketchengine.eu), English-Corpora.org (www.english-corpora.org) [formally corpus.byu.edu], and CQPWeb (cqpweb.lancs.ac.uk) offer many advantages for such work, as they not only circumvent the need for researchers to store datasets locally, but they are also designed to elegantly handle annotated data and offer fast access to very large datasets of millions and sometimes billions of words. On the other hand, online corpus portals can also restrict the type of research possible. One of the biggest limitations of most online portals is that they prevent researchers from uploading their own custom datasets (with SketchEngine being a notable exception). Another limitation is their need for datasets to be configured in a way that matches the database architecture on which the systems are built. This configuration can often be quite complex and can lead to the release of certain datasets being severely delayed or never released at all. One further limitation is the analytical tools available in online portals tend to be rather limited in scope. In contrast, offline software packages such as AntConc (Anthony, 2019), CasualConc (Imao, 2019), and WordSmith Tools

(Scott, 2019) allow for the direct upload of custom datasets and generally offer a wider range of analytical tools compared with those accessible through online corpus portals. However, they tend to also suffer from performance issues when handling heavily annotated small-scale datasets or very large datasets. The various weaknesses of both online and offline corpus tools suggest the need for a new type of corpus engine that can process both small-scale and very large-scale custom datasets that may be lightly or heavily annotated. In this paper, I propose a fast, scalable, and portable corpus database architecture that is designed to meet these needs. The database architecture is built using the HDF5 high-performance data software library together with a custom-designed index, which is similar to those used in web search engines.

In the paper, I will first explain the database architecture, before demonstrating its performance gains over traditional offline software engines. Next, I will show how the database architecture can handle a wide-variety of corpus datasets that vary both in size and degree of annotation. For example, the architecture can comfortably handle datasets such as the 10 m word Spoken BNC2014 (Love, 2019), which includes complex annotations describing speaker demographics, conversational turns, and non-verbal communication events. It can also comfortably handle much larger datasets, including the original 100 m word British National Corpus (BNC) (Bernard, 2007) and more recent, and much larger Web-based corpora. At the end of the paper, I will describe the steps needed to embed the database architecture into a standalone corpus tool in order for the architecture to be used for mainstream corpus research.

■ 第 1 セッションの概要

仁科 恭徳 (神戸学院大学)

2019 年 10 月 5 日(土)大会初日第 1 セッションでは、特定のディスコースのムーブ分析に関する発表が 2 件と、ソフト開発に関する発表が 1 件あった。第 1 セッションはいずれも英語による研究発表であったため、質疑応答時は英語による活発な意見交換が行われた。

第 1 発表者の Tatsuya Ishii 氏 (Hiroshima University) と Takeshi Kawamoto 氏 (Hiroshima University) は、基礎医学分野における研究論文の Result Section で用いられている 3 つのムーブに関して、文副詞に注目した分析結果を発表した。使用したコーパスは、2014 年刊行の 30

種のジャーナルに掲載された計 304 の論文から構成される約 160 万語の自作コーパスであった。Result Section では 3 つのムーブ (R1: Restating methodology, R2: Announcing results, R3: Commenting results) が使われており、各ムーブに特化した特徴的な文副詞として R1 では 4 種類、R2 では 12 種類、R3 では 8 種類が特定された。例えば、文副詞の *Interestingly* や *Notably* は R1 と R2 の繋ぎ目の役割を果たし、要約を示す言語マーカーとしての文副詞 *Together*; *Taken together*; *Collectively* は R2 と R3 の繋ぎ目の役割を果たしていた。全体として、とても分かりやすい発表であった。フロアからは、ムーブの特定方法に関する質問があり、発表者らは主観的にムーブを特定していたが、互いにダブルチェックすることで可能な限り客観性を担保しているという回答であった。他にも建設的なコメントや質問があった。

第 2 発表者の Yukie Kondo 氏 (Ritsumeikan University) は、日本の高級ホテルの公式ウェブサイトに掲載されている概要をコーパス化し、その談話構造をムーブの観点から分析した結果を報告した。分析対象となったホテルは、2018 年の Forbes Star Award Winners で 5 つ星、4 つ星、推奨ホテルの評価を獲得した計 29 のホテルである。比較のため、Kondo (2018) で構築・分析した計 47 のロンドンの 5 つ星ホテルのデータも用いられた。ホテルの概要の談話構造は、3 つのムーブ (Defining self, Establishing features, Establishing connections) と 3 つのステップ (History/architecture, Location, Facilities) からなり、その割合と特徴語に関してグレードの高いホテルと低いホテルとを比較することで、どのようにラグジュアリー・ストラテジーが前者で顕在化されているかが詳細に論じられた。分析の結果、ホスピタリティーの見せ方について日本とロンドンのホテル間に違いが見られ、ロンドンのホテルでは 'hotel-has-what' に、日本のホテルでは 'hotel/guests-do-what' に焦点が置かれていた。構造は類似しているものの、描写されている dream にも違いが見られた。全体として、発表者のプレゼンテーションスキルも高く、とてもリズムカルに発表が進んだ。フロアからは、今回の分析対象となったコーパスのサイズが小さいこともあり、どのようなコーパス分析メソッドを用いて分析したのかなど、多くの質問やコメントがあった。

第 3 発表者の Laurence Anthony 氏 (Waseda University) は、English-Corpora.org や CQPWeb といったオンライン・コーパス・ポータルや AntConc や WordSmith といった従来型のオフラ

イン・コーパス検索ソフトの弱点を補った、次世代型コーパス・データベース・アーキテクチャーに基づく AntConc の新バージョンを紹介した。検索処理がコーパスのサイズや複雑なアノテーションに左右されず動作が安定している。実際に、複雑なアノテーションが施されている 1 千万語の Spoken BNC2014 での検索などを披露した。タグ検索も可能であり、コンコーダンスライン上には、検索タグは表示されないなどの配慮がされていた。大規模コーパスの走りである BNC などにも快適に検索できるようだ。時間の都合上、用意されたスライドの全てを詳細に説明する時間はなかったこともあり、フロアからは使い方や検索に関する具体的な質問がいくつかあった。

第 1 日第 2 セッション

[司会] 菅原 崇 (岐阜工業高等専門学校)

「英語話し言葉コーパスにおける基本動詞 TAKE を含む連語表現 — 「意味のまとまり」に基づく体系的抽出方法の提案—

坂場 寛子 (東北大学大学院生)

岡田 毅 (東北大学)

本研究は、基本動詞を含む連語表現を体系的に抽出する手法を提案し、英語アカデミックスピーキングにおける基本動詞 TAKE の共起語と、その意味的特徴を明らかにすることを目的とする。連語表現には様々な定義と抽出方法がある。しかし、n-gram は必ずしも語連鎖と「意味のまとまり」の相関や、非連続的な表現を抽出できない。また、コロケーションはスパンを限定し、特定の品詞との共起に着目することが多いが、多様な語句と結びつく多義語の TAKE を調査対象とする場合、この方法では、TAKE を含む表現の中核の意味を担う共起語が調査範囲外となり、適切に特定できないこともある。

英語アカデミックスピーキングの代表的なコーパスとして、Michigan Corpus of Academic Spoken English (MICASE) を対象とし、以下 5 つの手順を経て TAKE を含む連語表現を分類する。

(1) レマとしての TAKE を MICASE のオンライン検索エンジンで検索し、TAKE を含むコンコーダンスラインを抽出

(2) TAKE+補部に着目 (Hunston & Francis, 2000) し、7 語を目安 (Lewis, 1997) として「意味のまとまり」(文中の TAKE を含む連語表現の意味理解に必要な語連鎖の最小単位) に区切る

(3) コンコーダンスライン 1 行につき、意味の中核を担う共起語を 1 語特定。(4) 以降の分析対象基準

は、頻度 5 回以上(Shin & Nation, 2007)

(4)特定した共起語の品詞と TAKE+共起語の意味を基準とし、TAKE+共起語を、句動詞、文法的コロケーション(不変化詞(句動詞として抽出した項目と接続詞を除く)との共起)、語彙的コロケーション、その他に分類

(5)English Vocabulary Profile を参考資料とし、連語表現を意味カテゴリに分類

(3)の条件を満たした TAKE+共起語(1,786 個)のうち、その他に分類した接続詞 and との共起(74 個)を除く TAKE+共起語(1,712 個)を連語表現として意味分類することができた。語彙的コロケーションの頻度が最多で、71 の異なり語との共起が確認された。句動詞は、調査対象とした全種類の使用が存在し、延べ 24 個の意味カテゴリに分類された。文法的コロケーションは、10 の異なり語と共起し、延べ 17 個の意味カテゴリに分類された。

本研究では、MICASE を分析対象とし、スパンの上限を「意味のまとまり」の範囲内という、従来より広範囲かつ柔軟性のある基準にし、調査する共起語を内容語のみならず機能語も含めることにより、これまでの限定的かつ機械的な手法では抽出できなかった、基本動詞 TAKE を含む多様な連語表現の抽出が可能になった。意味の観点からも当該連語表現を調査することにより、同じ語彙項目との共起でも複数の意味カテゴリに分類されることが明らかとなった。

「不可算抽象名詞の修飾と不定冠詞・ゼロ限定詞との関連性について～Bank of English のデータによる分析～」

小寺 正洋 (阪南大学)

不可算抽象名詞が不定冠詞(*a/an*)と共起する現象について、修飾語を伴うと生じるとの記述が多く、文法書や辞書に見られる。Swan (1995:139)は“we have to use *a/an* when we are limiting their meaning”とし、修飾語が *a/an* を強制するとの主張と考えられる。修飾語が *a/an* を誘発する強さについては、「必ず伴う」(Jespersen 1949, Swan 1995), 「しばしば伴う」(Swan 2005, 2016, MALED), 「時折伴う」(Sinclair 2017), 「伴うことが可能」(Berry 1993, Francis et al 1998, Downing & Locke 2002), 「個別性を強調しなければ *a/an* は不要」(Berry 1993:20-21), 「修飾語は *a/an* を強要しない」(Hewson 1972)など多様である。本研究では修飾語(形容詞, 関係詞節, 前置詞句)と限定詞(*a/an*, ゼロ限定詞, 他限定詞)の共起関係を、先行研究が挙げる 36 語について Bank of English のデータを基に調査し、修飾語が *a/an* を誘発

する力は弱く、関係詞節と前置詞句についてはゼロ限定詞(\emptyset)以外の限定詞を強く求める結果として *a/an* との共起が生じ、形容詞については *a/an* が形容詞を誘発すると考えるのが合理的であることを示す。

a/an と共起する不可算抽象名詞については「感情などの精神活動を表す名詞」(Quirk et al 1985, Swan 1995, 2005, 2016, Sinclair 1990, LDOCE, MEDAL)とするものがあるが、名詞のタイプについて詳述した研究は見当たらないため、調査対象は先行研究が挙げる以下の 36 語とした。名詞は NP の主要部に限定し *kind of N*, *N and N* などは除く。NODE 前後 10 語のどちらかが重複するものは削除。後述の理由で名詞の意味や解釈などは考慮しない。

anger, anxiety, aversion, awareness, coarseness, consciousness, courage, determination, dislike, distrust, eagerness, education, faith, familiarity, fierceness, fondness, happiness, hatred, help, kindness, knowledge, love, peace, pride, quietness, rationality, sadness, sensitivity, shyness, silence, sincerity, sleep, tranquility, understanding, unhappiness, warmth.

上記 36 語について(関係詞節と前置詞句では 22 語)、以下の 4 条件で名詞と限定詞の関係を調査した。

1. 直前に形容詞を伴う名詞。
2. 直後に関係詞節([N + *that* + V or Modal], [(Prep) + N + *which*], [N + V Past Participle])を伴う名詞。
3. 直後に前置詞句を伴う名詞。前置詞は、名詞が直後の前置詞句をその目的対象や志向対象とするもので、ヒット件数が最多のものを選択(e.g. *anger at his wife, knowledge of French, sensitivity to sunlight*)。
4. [Modifier + N], [N + Modifier]での *a/an* の生起率と [*a/an* + (0-2 words) + N]での修飾語の生起率とを比較し、修飾語と *a/an* との誘発関係を調査した。

前置修飾と後置修飾の重複については、関係詞節と前置詞句を伴う場合に形容詞の有無による 2 条件を設定したが、形容詞を伴う場合は後置修飾をすべて排除することが困難なため重複を考慮しない。

語の意味や解釈を考慮しない理由は以下の通り。不可算名詞と *a/an* との共起関係については、'a kind of'や'an instance of'の意味を表す場合(Sinclair 1992, Berry 1993, Biber et al. 1999, Downing & Locke 2002, CALD), コンテキストの具体性(Celce-Murcia & Larsen-Freeman 1999), 個別性の強調(Berry 1993), quantity の前景化(Hewson 1972:90-91), 有界性(Langacker 1987,

Radden & Dirven 2007)などにより説明されるが、いずれもコーパスデータによる客観的判断が難しい。抽象名詞では *temporal space* を主要領域とした場合、*episodic* か否かで可算性を判断するが、物理的有界性でなく解釈(construal)に依る場合が多い。例えば *There was Ø silence for a moment* の例を見ると物理的有界性(for a moment)が *a/an* との共起に直接影響を与えとは考えにくい。event より result を意味する場合の方が可算として受け入れられやすい(e.g. *invention*)と指摘するものもあるが(Payne & Huddleston 2002:337), event と result を客観的に区別するのが難しい。

「使役動詞 *make* の受動態補文に出現する原形不定詞とその語法」

村岡 宗一郎 (日本大学大学院生)

PDE において使役動詞 *make* は能動態補文には原形不定詞, 受動態補文には *to* 不定詞補文が出現するのか一般的である。

(1) PDE における使役動詞 *make* と不定詞

- a. They **made him sign** the contact against his will.(江川 (1991: 334))
- b. *Bert **made Jimmy to blush**. (Noonan (2007: 56))
- c. He **was made to sign** the contract his will. (江川 (1991: 335))
- d. *Peter **was made go**. (Gisborne (2010: 111))

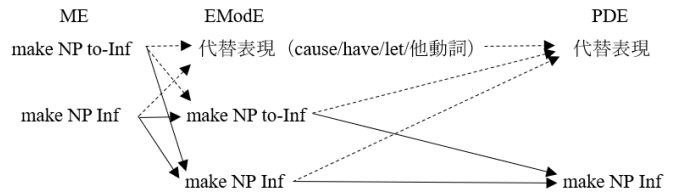
しかし、通時的に見ると態による不定詞の選択制限はなかったとされる。

(2) ME における使役動詞 *make* と不定詞

- a. she **maketh men mysdo** many score tumes. (PPI. Biii 122)(Mustanoja (1960: 533))
- b. þe veond hit **made me to don**. (Ancr. 136)(ibid.)
- c. Ich **am made reproce** up alle myn enemis, (PMPsalter, 30, 14) (松瀬 (1993: 6))
- d. [al thinges **ben**] **maked to dwelle** in present sight. (Usk TL.III.IV/167-168) (ibid.)

ME から PDE までの五つの英訳聖書を比較した村岡(2018)によれば、使役動詞 *make* は、借用語の流入およびその他使役動詞の出現とそれらとの意味の棲み分けにより EModE 以降 *make* が専ら強制使役を表す表現として用いられるにつれ、完結性・結果性のアスペクト特性をもつ原形不定詞を補文に取る構造が確立したとされている。その一方で、*to* 不定詞を補文にとる使役動詞 *make* は EModE 以降衰退し初め、LModE 末にはほぼ衰退したという。

(3) ME ~ PDE における使役動詞 *make* と意味変化 (cf. 村岡(2018))



一方使役動詞 *make* の受動態補文における *to* 不定詞の出現の理由については八木(1998)などにより明らかになってはいるものの、*be made Inf* がいつどのような過程を経て減少したのかについては明らかにはなっていない。本発表では、村岡(2018)で明らかになった能動態における原形不定詞補文の確立および *to* 不定詞補文の衰退の時期と比較し、*be made Inf* がいつどの様にして衰退していったのか、EEBO(Early English Book Online)および Google Book Corpora (British) を用いてその過程を明らかにしていく。本調査の結果から、使役動詞 *make* の受動態補文における原形不定詞の出現は 17 世紀後半にはじまり、19 世紀前半まで用いられていることから、*make NP to Inf* とほぼ同時期に衰退し、さらに本調査によって検出された *be made Inf* に出現する動詞の用例としては (4) に示すように *appear* や *believe* などの動作性の弱い動詞が多く検出されていることから、*be made Inf* は *causative* というよりもむしろ *resultative* として解釈されていたと提言する。

(4) 原形不定詞補文の用例

- a. it might easily **be made appare** in particular to the world, (EEBO, 1643)
- b. they **are made believe** that they are christians when they are not, (EEBO, 1658)

■第2セッションの概要

菅原 崇 (岐阜工業高等専門学校)

2019 年 10 月 5 日(土)大会初日第 2 セッションでは、特定の語法に関する共時的分析が 2 件、通時的分析が 1 件であった。

研究発表 1 は坂場寛子氏 (東北大学) (共著者: 岡田毅氏 (東北大学)) による「英語話言葉コーパスにおける基本動詞 TAKE を含む連語表現 — 「意味のまとまり」に基づく体系的抽出方法の提案—」, 本発表は基本動詞を含む連語表現を体系的に抽出する手法を提案し、英語アカデミックスピーキングにおける基本動詞 TAKE の共起語と、その意味的特徴を明らか

にすることを目的としたものであった。分析方法としては Michigan Corpus of Academic Spoken English (MICASE)を対象とし、TAKE+共起語を、「句動詞」「文法的コロケーション」「語彙的コロケーション」に分類、English Vocabulary Profile を参考資料とし、連語表現を意味カテゴリに分類、結果、TAKE+共起語(1,712 個)を連語表現として意味分類することができた。その中で語彙的コロケーションの頻度が最多で、71 の異なり語との共起が確認された。句動詞は、延べ 24 個の意味カテゴリに分類された。文法的コロケーションは、10 の異なり語と共起し、延べ 17 個の意味カテゴリに分類された。発表後は、オーディエンスより教育的応用として板場氏が TAKE を教える場合どのように学生らに伝えるのか、「句動詞」「文法的コロケーション」「語彙的コロケーション」の分類方法に曖昧性が残る、といった活発な質疑が展開された。

研究発表 2 は小寺正洋氏（阪南大学）による「不可算抽象名詞の修飾と不定冠詞・ゼロ限定詞との関連性について ~Bank of English のデータによる分析~」、本発表は修飾語(形容詞、関係詞節、前置詞句)と限定詞(a/an, ゼロ限定詞, 他限定詞)の共起関係を、先行研究が挙げる 36 語について Bank of English のデータを基に調査したものだ。使用コーパスは The Bank of English, 調査対象は先行研究が挙げる 36 語の抽象名詞, それらを①直前に形容詞を伴う名詞, ②直後に前置詞を伴う名詞, ③直後に関係詞節を伴う名詞, の 3 つに分類し, 修飾語と a/an との誘発関係を調査した。調査の結果, (1)Pre-/Post Modification + N が a/an と共起する確率は低い, (2)a/an+N は Adj との共起率が高い, N+PP は他の Det.との共起率が高い, N+RC は他の Det.との共起率が高い, といった結果が得られた。数量分析という点でデータが膨大であったため, オーディエンスとしては 20 分という短い発表時間で全貌を把握することが難しいものではあったものの, 質疑の際, 数値のみではなく語の意味にも着目したより深い調査を行う予定である, と発表者の将来的な展望を聞くことができた。

研究発表 3 は村岡宗一郎氏（日本大学大学院生）による「使役動詞 make の受動態補文に出現する原形不定詞の語法」, 本発表は発表者が 2018 年に明らかにした能動態における原形不定詞補文の確立および to 不定詞補文の衰退の時期と比較し, be made Inf がいつどの様にして衰退していったのか, EEBO(Early English Book Online)および Google Book Corpora (British) を

用いてその過程を明らかにしていく, というものであった。結果, 使役動詞 make の受動態補文における原形不定詞の出現は 17 世紀後半にはじまり, 19 世紀前半まで用いられていることから, make NP to Inf とほぼ同時期に衰退し, さらに本調査によって検出された be made Inf に出現する動詞の用例としては (appear や believe などの動作性の弱い動詞が多く検出されていることから, be made Inf は causative というよりもむしろ resultative として解釈されていたという提言がなされた。発表後は数量分析の妥当性, 例文の解釈などの点について指摘があったものの, 今後 make 以外の使役動詞についても同様の数量分析を行うことで考察の幅が広がるはず, との助言もあった。

第 1 日第 3 セッション

[司会] 杉森 直樹 (立命館大学)

Discourse functions of high-frequency phrase frames in argumentative essays: A learner corpus study

Joe Geluso (Iowa State University,
PhD Candidate)

Language is replete with formulaic language with some researchers estimating 50% or more being ‘formulaic’ (Erman & Warren, 2000; Nelson, 2018). Much research to date on formulaic language has focused on continuous multiword sequences of n length in the form of studies on n -grams and lexical bundles. More recently, in addition to continuous sequences of words, discontinuous multiword sequences, or ‘phrase frames’, have come into focus (Renouf & Sinclair, 1991; Römer, 2010; Gray & Biber, 2015). Phrase frames, or simply ‘frames’, are n -grams with an internal variable slot such as *the * of the* or *in the * of*. The asterisk represents a variable slot occupied by a ‘filler’ such as *end* or *beginning*. Previous work on frames, and even work primarily focused on lexical bundles, has reliably noted the high frequency and productivity of the frame *the * of the* as well as different permutations of frames rooted in *the * of*, such as *in the * of*, *at the * of the* (Biber, 2009; Gray & Biber, 2015; Garner, 2016; Nelson, 2018).

Given their high frequency, the present study attempts to elucidate the discourse functions of four or five-word frames (i.e., 4-frames or 5-frames) built around the core frame *the * of*. Specifically, high-frequency frames of the following structures were analyzed: 1) *the * of the*, 2) *preposition + the * of*, and 3) *preposition + the * of the*. Therefore, all

target frames in this study were 4-or 5-frames. The target frames were analyzed as they appear in English argumentative essays authored by native speakers of Japanese, Spanish, and English. Texts were drawn from the Japanese and Spanish sub-corpora of the International Corpus of Learner English (ICLE) and the LOCNESS corpus (Granger et al. 2009). Each text was coded by the researcher for the topic of the essay (something not included in the original corpus files), and then rated for proficiency level by two trained raters. Final proficiency scores for each text were generated via a Many-Facet Rasch Measurement. Frames were then extracted using a custom Python script that captured the target frames along with text that preceded and followed the frames to create concordances and stored in an Excel sheet. The concordances were then manually analyzed to determine the discourse function of each instance of each frame. The functional taxonomy introduced by Biber et al. (2004) was used to code for discourse function. Primary functions that the frames achieved were referential, stance, and text organizing functions. Furthermore, frames were coded for sub-functions within the three higher-level functions. Ultimately, 3,693 target 4 and 5-frames were coded for discourse function.

After coding for discourse function, a multinomial logistic regression was conducted to ascertain the likelihood that a given frame would be used for a given discourse function. For example, how likely is it that the frame *the * of the* or *at the * of* will be used for the sub-function of marking location or time within the referential function? In order to avoid the assumption that all differences between texts stem from differences in first language (L1), four independent predictor variables were added to the model as blocks: 1) L1, 2) topic, 3) proficiency score, and 4) the preposition such as *in* or *at*, or no preposition (e.g., *the * of the*). Broadly, results show that the L1 English and Spanish speakers use the target frames much more frequently than the L1 Japanese speakers. Furthermore, while each predictor reveals significant effects, the models that included proficiency scores and prepositions as predictors performed significantly better than those that did not.

Exploring the impact of the Brexit referendum on the encyclopedic knowledge of English speakers: A Latent Dirichlet Allocation analysis

Naoki Kiyama (The University of Kitakyushu)
Yoshikata Shibuya (Kanazawa University)

In this paper, we explore, from a corpus-based cognitive linguistic perspective, issues of how and to what extent historically pivotal events can affect the meaning of words. The event that we take on as a case study is what is widely known as Brexit (i.e. British exit from the European Union). The so-called United Kingdom European Union membership referendum was held on 23 June 2016. The result was that the United Kingdom (UK) will leave the European Union (EU). The news made headlines around the world, resulting in a number of upheavals in many domains, including the financial markets' immediate negative reaction and the announcement of the resignation of the then Prime Minister David Cameron.

The present study investigates the impact brought about by the Brexit referendum on the minds of English speakers. Since the consequences of Brexit are, needless to say, far-reaching, it is not our intention to provide a comprehensive account in this paper. Instead, here we pay special attention to two keywords that are very strongly associated with the Brexit referendum. One is the very name of the event itself, *Brexit*, and the other is the word *referendum*. Our research question is how and to what extent the Brexit referendum affected the meaning and/or perception of the words *Brexit* and *referendum*. To address the issue, we apply a generative statistical model called the Latent Dirichlet Allocation (LDA; Blei et al. 2003). LDA is a probabilistic topic model that allows one to construct statistically meaningful clusters of words out of texts in question. We applied LDA to the data retrieved from the NOW corpus (News on the Web; Davies 2013).

The present study draws on the view of meaning advocated in cognitive linguistics. One of the fundamental assumptions of cognitive linguistics is that "we have to call on our encyclopedic knowledge in order to properly understand a concept" (Croft and Cruse 2004: 30). The present paper is especially concerned with the dynamic nature of our encyclopedic knowledge. Our encyclopedic knowledge is in constant flux, making adjustments to respond to any changes, be it social or cultural. The goal of our study is hence to capture how our

encyclopedic knowledge recruited for understanding the words *Brexit* and *referendum* transitioned before and after the Brexit referendum.

We describe a series of significant topics identified by LDA (Figures 1 and 2 illustrate a subset). The results capture our intuition that the topics that are highly relevant to the campaign for and/or discussion on the Brexit referendum itself were extremely prominent for a month or so before the voting, while topics in which negative aspects of Brexit are the subject for discussion became significantly marked after the referendum. The results show that the encyclopedic knowledge accessed by speakers to understand the words *Brexit* and *referendum* drastically changed before and after the Brexit referendum, arguably into a more negative direction. The cognitive linguistic approach to encyclopedic knowledge has typically been qualitative in nature. However, given the centrality of our encyclopedic knowledge in understanding concepts and also its inherently dynamic nature, a corpus-based diachronic approach should be actively taken. The present study is an attempt in that direction.

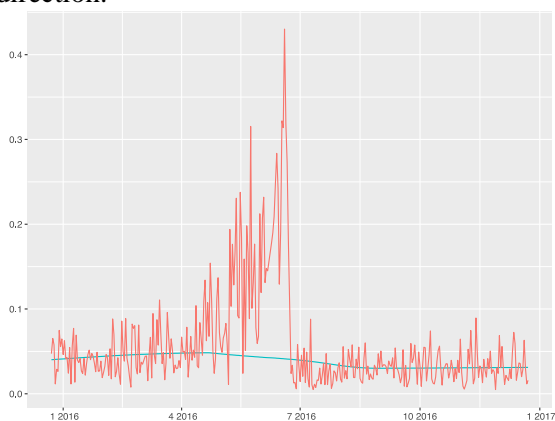


Figure 1: A topic identified via LDA for the word *Brexit* (label: “Campaign for the Brexit referendum”)

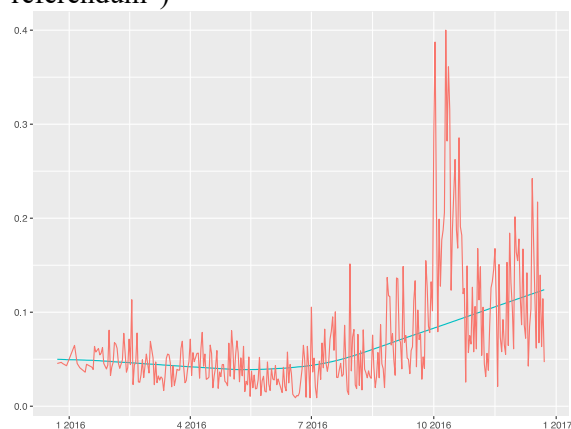


Figure 2: A topic identified via LDA for the word *referendum* (label: “Economic uncertainty associated with the referendum results”)

■ 第3セッションの概要

杉森 直樹 (立命館大学)

「Discourse functions of high-frequency phrase frames in argumentative essays: A learner corpus study」の発表後の質疑応答では、日本人英語学習者は必要な冠詞を落としてしまう誤用が多いとされているが、それをどの様に処理したのか、またそれが結果にどの程度影響したのかについての質問や、フレーズのスロットの中には実際にはどのような語が入っていたのか、今回の研究結果が英語教育にどのようなことを示唆するのか、などの質問がなされた。

「Exploring the impact of the Brexit referendum on the encyclopedic knowledge of English speakers: A Latent Dirichlet Allocation analysis」の質疑応答では、referendum 以前は Brexit がもたらす利益と不利益の両方が語られていたが、referendum の後では敗北した反対派の不利益に関する意見だけが紙面に掲載される傾向にあったのではないかというコメントがなされた。また、referendum の前後でグラフが変動しているトピックが他にもある中でなぜ特定のものだけを選んだのか? という質問や、discourse を考慮すると機能語についても見ないといけないのではないかというコメント、世代間で referendum の投票傾向の違いはあったかについての質問などがなされた。

本セッションには McEnery 氏も参加され、それぞれの研究発表についての的確な質問や示唆に富むコメントをされていたことが印象的であった。発表者にとっては大変貴重な研究上のフィードバックを得る機会になったと思われる。

第1日第4セッション

[司会] 内田 諭 (九州大学)

「語彙頻度の圧縮による作品区分・文体変異の分析確度向上—クラスター分析—」

後藤 克己 (中部大学大学院生)

クラスター分析を用いた著者推定では、作品間の距離をユークリッド法で測ることが多い。その距離の基礎となる語彙の頻度差は一般に高頻度語ほど大きく、また作品には頻度上位語が際立って多く生起する特徴がある。そのためクラスタリングの大要は、機能語を多く含む頻度上位語で形成され、多くの内容語の生起特性は反映されにくい。従ってこの手法は、同じ作者の作品に一貫して高頻度で生起する語に着目す

る著者推定には有効だが、同一作家の作品区分や、作品内の文体変異の分析には効果的とは言えない(図1, 2)。

本研究はそのような分析において区分確度を高める手法を探るため、テストコーパスとした Dickens, Thackeray, 及び G. Eliot の各 2 作品(計 6 作品)を、それぞれ 10,000 語程度に分割(計 62 セクション)し、セクション間距離にユークリッド法以外の定義(マンハッタン, IRad(情報半径)等)を適用したデータ、あるいは頻度自体を平方根, 立方根等で圧縮したデータを用いてクラスター分析し、得られた区分性—各クラスターに含まれるセクションの作家/作品不純度の大小—を Gini 係数で評価したものである。結果は、距離定義では IRad による場合に、また、頻度圧縮では平方根と立方根の間の圧縮度となる"2.5 乗根"圧縮した場合に最大の区分性が得られ、特に後者の場合、62 のセクションは上位 500 語以上で殆ど完全に 6 区分された(図3, 4)。これは頻度上位語だけでなく、広範な頻度ランクの語の生起特性がクラスターリングに寄与した結果であり、作品区分、更には作品内の文体変異分析に有効な手法になり得るものと考えられる。

質疑では、単語に加え bigram や trigram 等による分析、2.5 乗根の圧縮度の一般性、分析に用いた 6 作品の選定基準等についてご質問/コメントをいただいた。

6 クラスター中 4 クラスターに異作品が混在

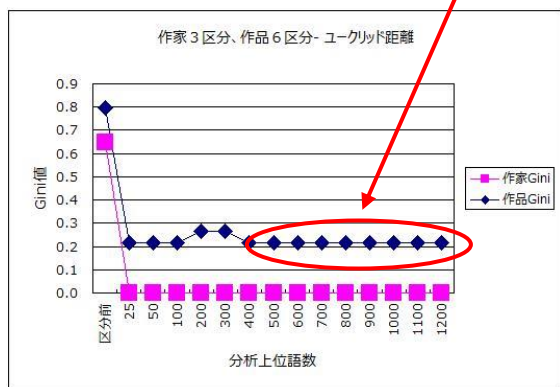


図1 ユークリッド距離(euc)による不純度

不純な4クラスターはサブクラスターでも不純

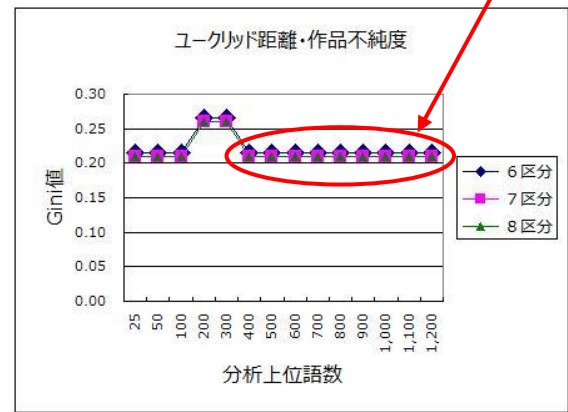


図2 区分数による不純度の変化 (euc)

6 クラスターは完全に区分(異作品の混在なし)

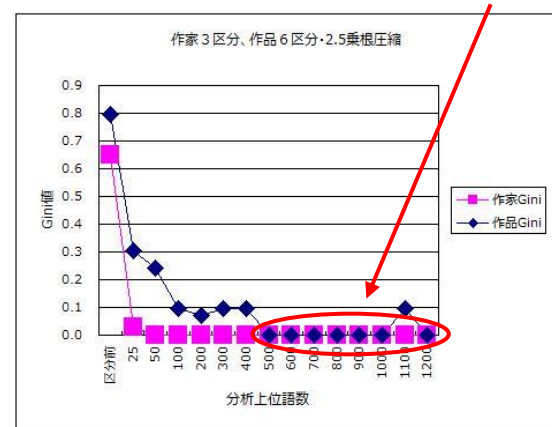


図3 2.5乗根圧縮による不純度

7区分では300語以上で完全に区分

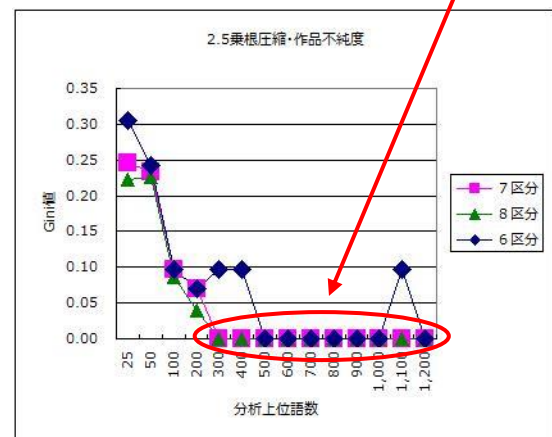


図4 区分数による不純度の変化 (2.5乗根)

「Helsinki Corpus ME3-EModE3 での binomials の 史的変遷—特に頻度と語源から」

谷 明信 (兵庫教育大学)

本研究は、近年再び注目されるようになった time and tide のような A and B 形式の binomials

(以下, BN) について, Helsinki Corpus の ME3, ME4, EModE1, EModE2, EModE3 (1350-1420/1420-1500/1500-1570/1570-1640/1640-1710) を対象に, 時代とテキストタイプの分類に基づき, A and B 形式の BN の頻度と語源を調査し, その史的変遷を検討した。さらに, BN の繰り返し度についての調査結果も報告した。その際, 英語の語彙の変遷との関連についても考察し, 議論の中では Helsinki Corpus の問題点についても触れた。

収集した type 5801, token 8257 の BN の用例調査から得られた主要な結果は, 以下の通りである。

(1) テキストタイプにより, BN の頻度と語源構成が大きく異なる

(2) 全体では BN の正規化頻度が ME4 で一番高くなり, EModE1-3 にかけて漸次低下していく

(2a) ME4 での正規化頻度が上昇した要因は, 名詞から構成される BN の増加である

(3) テキストタイプ中, 平均では LAW で正規化頻度が一番高く, LET NON- PRIV (公式書簡) がそれに続く

(3a) LET NON-PRIV は ME3-4 での高頻度が EModE1 以降は半減する (理由については不明)

(3b) テキストタイプについては, 正規化頻度が科学系文書では低く, また喜劇でも低い

(4) 語源の点では, 全体的には, ME3-4 では OE (古英語由来の語) +OE の BN の方が OF (古仏語由来の語) +OF の BN より割合が高いが, EModE1 からは逆転し, OE + OE の BN の割合は減り続け, OF + OF の BN の割合は増え続ける (なお, OE+OF と OF+OE の割合は時代を通じて変動が比較的少ない)

(4a) 全体として, OF+OF の BN の割合が高いテキストでは OE+OE の割合が低く, 逆に, OE+OE の BN の割合が高いテキストタイプでは OF+OF の割合が低い;

(4b) 全体的な傾向に反して, LAW では ME4 から EModE3 にかけて, OF + OF の BN の割合が減少し, 一方 OE + OE の BN の割合が増加する。これはこのテキストタイプの自国語化 (vernacularization) のためと推測した;

(4c) 語源の変遷を品詞別で見ると, 名詞でだけ ME4 において OF+OF の BN が OE+OE よりも割合が高くなり, EModE1 で変化する動詞と形容詞に先んじている

(4d) 副詞から構成される BN のみが, OE+OE の BN の割合が高く, 他の品詞の BN と大きく異なる;

(5) 繰り返し度については, 時代による変動は少ないが, 全体としては BN の頻度自体が減少

する EModE1 以降に, 繰り返し度も減少する;

質疑応答では, 韻文を散文と一緒に調査する問題, 公式書簡と比べ史的書簡での BN の頻度の低さは書簡自体の長さの問題の影響ではないか, また, 特定の binomials の変異などについての質問がなされた。

■第4セッションの概要

内田 諭 (九州大学)

本セッションの第1発表は, 英語文学作品コーパスを対象にクラスター分析を実施するための最適なパラメータの探索を行ったものである。文学作品に出現する単語の頻度を入力として, 作家および作品の分類を実施する際に, 頻度の 2.5 乗根が有効であることなどが示された。この手法の一般性を高めるため, 今後, 他の作品での検証や単語の頻度以外を入力とした場合の検証などが期待される。

第2発表は, 英語の A and B の形式 (binomials) の変遷を, 通時コーパスを使って読み解く試みである。テキストタイプや語源, 品詞の違いなどによる考察が示され, 通時的な変化が明らかとなった。また, 通時コーパスにおける問題点も指摘され, コーパスを使った語彙・表現の変遷の研究について示唆に富む発表であった。今後, 後期近代英語～現代英語との対照的な研究などが期待される。

第2日第5セッション

[司会] 木山 直毅 (北九州市立大学)

「この上なく…だ」を示す as … as 構文と about との共起傾向を探る—(about) as … as it gets, (about) as … as they come, (about) as … as you (can) get—

松田 佑治 (立命館大学)

・問題の所在

「この上なく…だ」を as … as 構文で示すと, 概ね以下のように表現できる。

- (1) a. John is (about) **as happy as can be.**
- b. John is (about) **as happy as he can be.**
- c. John is (about) **as happy as a boy can be.**
- d. John is (about) **as happy as happy/*happiness can be.**
- e. John is (about) **as happy as it gets.**
- f. John is (about) **as happy as happy/*happiness gets.**
- g. John is (about) **as happy as they come.**
- h. John is (about) **as happy as you (can) get.**

i. John is (about) **as happy as it's possible to be.**

太字箇所は概ね very/extremely と言い換えができ、それぞれが happy を強調している。(1a-i)の最初の as の直前には, probably/possibly/nearly/almost よりも, 話者の主観的判断・価値判断において「この上なく...だ」という断定を和らげる hedge の役割を持つ about が共起しやすいように思われる。発表者が知る限り, その点を指摘している先行研究はない。そこで本発表では, 正規表現を用いて, (1a-i)のような as ... as 構文と about が共起しやすいのかどうかを検証する。

・分析方法

対象とするコーパスは COCA full text である。紙幅の制約上, (1h)型のみ示す。

```
正規表現 : perl -ne 'while
(∕b(is|are|am|was|were|be|being|been) ([a-z-]+)
as( [a-z-]+){1,3} as you( can|could)? ge∕b/gi) {$a =
lc $2; print "$a\n"}' COCA full text | sort | uniq -c |
sort -rn | head -5
```

意味: be (...) as ... as you (can) get の be 動詞と最初の as の間の 1 語の頻度表 (上位 5 位まで)。
as ... as の...は 3 語までに制限する。

結果: 1. about (53) 2. already (2) 3. still (1)
 4. probably (1) 5. not (1)

()内は頻度数。また, be (...) as ... as you (can) get の COCA full text における全体頻度は 117 である。

・結果と考察

本調査の結果, (1e-h)型では about が共起しやすいという主張が支持されると思われる。しかし, (1a-d)型に関しては, 非該当例を多く含んでしまうことから, 現時点では about が共起しやすいという結論までには至っていない。今後の課題として, (1a-d)型を対象とした, より精緻な調査方法が求められる。

「wisely の生起位置と意味機能について」

西村 知修 (西南学院大学大学院生)

本研究の目的は, 副詞 wisely の生起位置と意味機能をコーパスを用いて調査することである。wisely や carefully などの副詞は, 大まかに以下の 3 つの位置に生起する可能性があり, 位置によって異なる機能を持つ。

(1) (Wisely,) John (wisely) dropped the vase

(wisely).

動詞後方の位置では様態副詞として, 文頭の位置では話者の価値判断を表わす主語指向副詞として機能し, 中間位置ではどちらの機能が曖昧であるとされる(Jackendoff, 1972)。様態性が低い副詞の場合, 中間位置では価値判断の機能しかないという指摘もある(Huddleston & Pullum, 2002)。このように様々な提案があるが, コーパスを用いてその 2 つの機能を同時に持つ副詞を詳細に検討したものは少ない。2 つの機能を持つ副詞の代表格として頻繁に挙げられる wisely を調査することによって, 先行研究の妥当性を検討する。

上述の目的のため, COCA にある 2401 例の wisely を取り出し, それぞれの位置における生起数, 共起する動詞, 使われている文脈などを調査した。wisely が不定詞節や動名詞節の内部, 受動文に生起する例は, 分離不定詞などの要因が絡み複雑になるため分析対象外とし, 残った 1524 例を分析した。

分析の結果, wisely が文頭に生起する例は全体の約 6%, 中間位置は約 48%, 動詞後方位置は約 46%であった。中間位置の wisely と共に現れる動詞上位 15 位までを示すと decide, choose, keep, avoid, say, take, point, leave, note, put, refrain, refuse, hold, use, suggest, turn, counsel となった。動詞後方位置では use, choose, spend, invest, act, nod, splurge, pick, say, shop, rule, speak, make, manage, wield となった。いくつか共通する動詞もあるが, 2 つの位置の wisely が好む動詞には違いがあることが明らかになった。これは 2 つの位置の wisely が異なる機能を持つ証拠となる。文頭位置に関しては生起数が少なく比較が困難だが, 中間位置の wisely と共起する動詞と共通するものが幾分多いようである。

choose のように中間位置・後方位置ともに生起数が多い共起動詞が存在することは, 中間位置の wisely が機能の曖昧性を持っていることを示唆しているように思われるが, 詳細に検討すると異なる結論が導かれる。例えば wisely が choose の後方で使用される場合, choose は自動詞の割合が約 40%, 他動詞の割合が約 60%で, 目的語として名詞をとる場合がほとんどであり, また命令文で生じる割合が約 40%になる。一方, wisely が choose の前の中間位置に生じる場合, choose は全て他動詞で不定詞節を目的語にとる割合が約 70%になり, 命令文には 1 例も生じない。これらの差異の原因の多くは, 価値判断と様態修飾の機能の違いに帰結することが

できる。よって2つの位置の *wisely* に共通して多く現れる共起動詞は、中間位置の機能が曖昧であることを示すのではなく、むしろ位置による機能分担が明確に存在する証拠であり、様態性が低い副詞の場合、中間位置における機能の曖昧性はないという Huddleston & Pullum (2002) の主張を裏付けるものとなる。

■第5セッションの概要

木山 直毅 (北九州市立大学)

本セッションでは英語文法記述に関する2件の発表があった。どの発表でも建設的な質疑応答が行われる内容であった。以下、発表毎の質疑応答を完結にまとめる。

第1発表者の松田佑治氏 (立命館大学) は、COCA full text を用いて「この上なく...」の意味を示す *as ... as* 構文を調査した。発表では本構文において *about* と共起する傾向が高いことが指摘され、その理由として、「この上なく...」という断定を避けたいという話し手の意識がより強く働くためではないか、という見解が示された。またそれにより *about* を独自のクラスとみなす必要性が論じられ、辞書に対し新たな項目として本調査の内容を取り入れる可能性について示唆された。質疑応答では、本構文と共起する *about* と通常の *as ... as* 構文と共起する *about* との違いや、辞書的記述にどのような形で貢献するのか、などの質問が寄せられた。

第2発表者の西村知修氏 (西南学院大学[院]) は、*wisely* が生起する位置による意味や語用論的意味の差異を調査した。氏が挙げた先行研究では *wisely* の位置が動詞の直前に生じれば価値判断を修飾し、動詞の直後に生じる場合は様態及び価値判断の修飾が可能であり、曖昧であるとされている。この先行研究に対し、西村氏は COCA (オンライン版) を用いて調査し、*refuse* や *avoid* といった拒否を表す動詞クラス等は *wisely* が動詞の前に現れること傾向が強く価値判断の修飾を好むが、様態修飾は取らない、あるいはその逆の傾向を示した。この結果より、本発表では動詞の直前に *wisely* が現れる場合、特定の動詞 (クラス) と現れると様態修飾である傾向が強く、具体性の高いレベルにおいて動詞の直前に現れる *wisely* の曖昧性がなくなるという主張がなされた。質疑応答では構文と *wisely* の関係について、情報構造の観点からのコメントが寄せられた。

第2日第6セッション

[司会] 能登原 祥之 (同志社大学)

「大規模コーパスを用いた CEFR レベル別の英語コロケーションおよびフレーズの抽出」

内田 諭 (九州大学)

南里 豪志 (九州大学)

英語学習における語彙の習得の重要性は言を俟たないが、近年、単語と単語の組み合わせであるコロケーションの習得の重要性が多く研究者によって指摘されてきた (Sinclair, 1991, Lewis, 2000, 堀, 2011)。その一方で、学習上の困難さも指摘されており、上級レベルの学習者であっても基本語のコロケーション (共起語) の習得は難しい (cf. Altenberg and Granger 2001)。その一因は、学習者が学ぶべき共起語がレベル別に明らかにされていないことにあると考えられる。English Vocabulary Profile や CEFR-J Wordlist など、語彙のレベル別リストは存在するが、コロケーションのリスト決定版と言えるものは管見の限り見当たらない。また、近年学習者向けのコロケーション辞書が出版されているが、多くの場合、共起語がリストされているだけで実際にどのように使われるかという情報に関しては限られた用例に頼るしかなく、学習者にとって十分な情報であるとは言い難い。

本研究では、CEFR-J Wordlist に含まれる内容語を中心語とし (action_N(A1)など)、同リストに含まれる別の内容語 (take_V(A1), appropriate_J (A2)など) が共起関係にある場合のリストを iWeb Corpus のデータを用いて CEFR レベルごとに作成した。さらに、中心語と共起語の両方を含む n-gram を品詞パターンに集計し、フレーズを抽出した。iWeb Corpus は約 625GB に及ぶ巨大なデータであり、これを利用することで比較的長いフレーズを十分な頻度で抽出することが可能となる。しかしながら、一般的なコンピュータでは処理が難しいデータサイズであるため、スーパーコンピュータを用いて計算を行った。スーパーコンピュータは多数の CPU コアを搭載しているため、共起関係の解析や、n-gram の抽出の処理を中心語ごとに分割し、それらを並列処理することで、大幅な高速化を実現できた。

role (A1)を中心語として抽出結果の具体例を示す。まず、この語の共起語として key_J(A1), play_V(A1)などが抽出される。次に中心語(role)と共起語 (ここでは key) を含む 2 gram~6 gram を抽出し、中心語と共起語以外は品詞情

報だけを残しマスクして集計する (*_A key_J role_N *_I, _A key_J role_N *_I *_V など)。一定の頻度を閾値とし、それぞれのパターンの具体例を集計する。例えば、A key_J role_N *_I *_V の場合は、a_A key_J role_N in_I help_V (1415), a_A key_J role_N in_I develop_V (511)などのフレーズが抽出された。このような処理を CEFR-J Wordlist の全ての内容語について行った。このアプローチの利点は、第 1 に学習者が中心語の使用方法を具体的に知ることができ、特に発信目的で利用できるリストになるということである。「語」から「フレーズ」への連続性が可視化されるため、使用法のイメージが容易になる。第 2 に、n gram をフレーズパターンとして集計を行っているため、単語と単語の集計では拾いにくい句動詞を含むパターン (go_V against_I medical_J advice_N) や前置詞などからなる慣用パターン (range_V from_I high_J to_I low_J) も含むことができることである。第 3 に、抽出されたフレーズは CEFR レベルに基づく重要語が必ず複数含まれており、学習価値の高いものとなっていることが挙げられる。

「特定の文法項目抽出のためのライティング課題の設計：学習者コーパス構築への示唆」

工藤 洋路 (玉川大学)
内田 諭 (九州大学)

学習者コーパスを構築する際、特定のライティング課題が与えられ、自由作文を課すことが多い。例えば、ICNALE (International Corpus Network of Asian Learners of English) は、It is important for college students to have a part time job.などの陳述に対して賛成か反対かを書く課題を使って構築されている (Ishikawa, 2011)。このような課題設定は自由に記述できるため多様な学習者の中間言語を観察できるという利点がある一方、特定の文法事項の使用不使用をコントロールすることは難しく、学習者の時系列的な文法事項の発達を観察することが難しい。そこで工藤・内田(2019)では、ライティング課題において、指示文の内容や詳細さが学習者の文法産出状況にどのような影響を与えるかを調査し、適切な指示文を与えれば特定の文法事項を抽出できる割合が変化する場合があることが判明した。例えば、「あなたが行ったことのある場所について、3 つの文を書いてみよう」という課題と、「あなたが今までに複数回行ったことのある場所 [飲食店や観光地など] について、回数を含めて 3 つの文を書いてみましょう」という課題を比べた場合、現在完了形を使

用した学習者は、前者の課題では約 35%であったのに対し、後者では 90%近くであった。このことは、ライティングの学習者コーパスを構築する際には、蓄積していく言語データをどのような課題を用いて抽出すべきかに留意する必要性があることを示唆している。ただし、工藤・内田(2019)は、当該の文法事項を学習してから一定の期間が過ぎている大学生の作文データを分析したものであることから、すべての学習者について、課題の指示文の変化が特定の文法事項の使用率に変化を及ぼすと言えるかは明らかにはならなかった。そこで、本研究では、当該の文法事項を学習してからそれほど長い時間が経過していない中学 3 年生約 100 名を参加者として、大学生と同様のライティング課題を実施し、その結果を大学生のデータと比較することで、能力や学習後の期間の長さの違いによって、特定の文法事項の使用率がどう変化するかを調査した。具体的には、参加者を 2 つのグループに分け、1 つ目のグループには、(1) 教科書の活動からモデル文を取り除いた課題を実施し、2 つ目のグループには、(2) 教科書の活動からモデル文を取り除き、さらに指示文を書き直した課題を実施した。中学生の作文データと大学生のデータの比較を行った結果、中学 3 年生についても、大学生と同様に、当該文法事項の使用率が指示文によって異なり、また、抽出しやすい文法事項とそうでない文法事項があることが分かった。一方で、中学生の場合、特に現在完了形の誘出を意図した課題では、(1)・(2)の違いに関わらず非常に高い使用率を示した。この要因は、現在完了形が授業において「習いたて」だったことが考えられる。この研究により、指示文の詳細さなど、ライティング課題の設計方法によって、テーマやトピックが同じであっても、学習者が特定の文法事項を使用できる確率が変わり、さらに、学習段階および学習時期の違いによっても、その確率が変化することが明らかとなった。また、ICNALE での出現割合との比較から、特に現在完了はライティング課題の設計段階で意図しないと引き出すことが難しいことが示唆された。これらの結果を踏まえると、学習者コーパス構築のライティングの課題の設計段階で特定の文法事項をある程度コントロールできることが示唆される。これにより、特定の文法事項をターゲットとした学習者コーパスの設計が実現できる可能性がある。一方、課題設定のみでは抽出できない事項もあることが示唆され、将来的な課題である。

■第6セッションの概要

能登原 祥之 (同志社大学)

第6セッションでは、大規模コーパス iWeb Corpus を用いたコロケーションおよびフレーズ研究1件と、学習者コーパスの収集に関わるライティングタスクの設計研究 (特に指示文が学習者の英文に与える影響) 1件、の計2件の発表があった。

まず、第一発表者の内田諭 (九州大学) ・南里豪志 (九州大学) 各氏からは、iWeb Corpus (約140億語) のデータをスーパーコンピュータシステムで高速に処理したコロケーションおよびフレーズ調査 (CEFR-J Wordlist の内容語に注目) の手順とその結果が報告された。発表では、特に、共起語リストの作成法、コロケーションおよびフレーズの抽出法、また、抽出されたリストの教育的意義、がそれぞれ丁寧に説明された。質疑応答では、このサイズのデータ分析にスーパーコンピュータが必要なのか、データベースはどのような仕組みか、実際のデータ検索はどのような仕組みか、など、スーパーコンピュータで処理されるデータとその検索プロセスに関心が集まり、活発な意見交換がなされた。

次に、第二発表者の工藤洋路 (玉川大学) ・内田諭 (九州大学) 各氏からは、昨年度に続き、3種の文法項目 (現在完了, *to* 不定詞, 受動態) を自然に抽出できるライティングタスクの要件 (特に指示文) の研究成果が報告された。今回は、英語の熟達度レベルの違い (中学3年生と大学生) と文法項目の学習時期の近さ (現在完了と *to* 不定詞の2項目は中学3年生にとって最近学習したもの) により英文が異なるかが研究課題とされた。発表では、高校の英語教科書 *Vision Quest* のタスクをもとに (1) 原文の場合 (e.g., あなたが行ったことのある場所について、3つの文を書いてみよう) と (2) 改訂文の場合 (e.g., あなたが今までに複数回行ったことがある場所 [飲食店や観光地など] について、回数を含めて3つの文を書いてみよう) とで比較した調査結果とその解釈が詳しく報告された。質疑応答では、最近学習した定型表現 (e.g., *I've been to*) が出てこないよう指示文を工夫できないか、どのような意図で今回の文言を追記したのか (特に *to* 不定詞の使用を期待し「心がけて」と添える場合がわかりにくい)、改訂文で追記したどの文言が効いて意図した文法項目が使われたのか、など、指示文と学習者の

英文との関係の探る質問が出て丁寧な意見交換がなされた。

本セッションを通して、想像を絶する大規模のコーパスデータをより高速に調査できる時代が来ていると素直に感動を覚えた。一方で、タスクに基づくパフォーマンス (task-based performance) は状況に左右され複雑であるため、データを分析・解釈する際、表現に隠れている多様な背景 (言語使用者の特徴やデータ収集のタスク環境など) の違いを丁寧に読み解くことも重要であると改めて感じた。

■講演 (要旨の再掲)

Revisiting A Corpus—the BNC 2014

Tony McEnery (Lancaster University)

Major corpora, such as the British National Corpus, age. As they grow older their value changes, but is not lost. Instead of being representative of contemporary speech and writing, they become precious time capsules recording language as it was at some point in the past. That makes the task of re-creating such resources for the present a pressing one. Yet it is also a complicated undertaking. To demonstrate this, I will review the construction of the spoken and written BNC 2014. This corpus has been built to provide an up to date record of present day British English, broadly comparable to the BNC produced in 1994.

Recreating such a corpus may sound like a trivial task – just build the corpus again but draw material from the present. However, time changes not only a corpus, it also changes the context of collection. This means that the corpus builder has choices to make. In this talk I will outline the many decision points that we had to face when building the BNC 2014. All of these decision points are ones at which a choice must be made, and all of the choices made make the BNC 2014 slightly different to the BNC 1994. Some of the decision points arose from a realization that the decisions made in the construction, or execution, of the BNC 1994 were either not ideal or simply wrong. Others were faced when we considered the different legal and ethical environment in which the BNC 2014 was collected relative to that which existed when the BNC 1994 was created. Yet others are forced upon us by language change – the context within which language is produced, and the form it takes, has changed from 1994. Finally, we also have to accept practical limitations, such as the availability of cash

and resource in corpus construction, and face up to the decisions that flow from that.

Overall I will argue in this talk that as well as it not being possible to simply recreate some major historic corpora, it is also not desirable to do so. What is desirable, I will argue, is to document, honestly and thoughtfully, the decisions that were made in corpus construction so that users are aware of where differences exist and can take them into account in their analyses

■シンポジウム（要旨の再掲）

Gazing into a crystal ball: what you can see in the future of corpus linguistics

Yukio Tono (Tokyo University of Foreign Studies)

Shin'ichiro Ishikawa (Kobe University)

Hitoshi Isahara (Toyohashi University of Technology)

Tony McEnery (Lancaster University)

This symposium aims to discuss the future of corpus linguistics in this changing world. Since the beginning of an electronic corpus such as the Brown Corpus in the early 1960s, corpus linguistics has gone through various theoretical and methodological shifts. Theoretically speaking, corpora are designed to be representative of a particular genre or language variety. Recently, however, the sheer size of data, mainly collected from the web, seems to override the concept of representativeness. Methodologically speaking, corpora are now increasingly used by researchers in other disciplines of linguistics and social sciences. People are more familiar with the term such as "AI" or "big data" now, and the notion of corpus linguistics as an independent discipline seems to be less visible these days.

In this symposium, each of the four speakers will reflect on the past and present of corpus linguistics from their perspective. They hope to answer questions such as "What improvement can we make regarding corpus design and annotations?", "Can we see new types of corpora?", "Will corpus linguistics stand as an independent discipline of linguistics?", "What kind of new use will corpora be put to in other areas of sciences?", among others.

■ワークショップ（要旨の再掲）

「人工知能とデータ — 自然言語処理とコーパスを例に」

井佐原 均（豊橋技術科学大学）

将棋や碁などで人間に勝つようになった人工

知能は人間の味方でしょうか、敵でしょうか？言葉の処理においても、2016 年末に Microsoft や Google がニューラル機械翻訳という新しい人工知能型の自動翻訳システムでのサービスを開始し、人間と同じ程度の翻訳が可能となっています。人工知能システムは大量のデータを用いた機械学習が基盤となっています。ニューラル機械翻訳でも億単位の文が学習に使われます。このワークショップでは、人工知能型翻訳システムの現状をその長所と弱点を示しつつ解説します。また、人工知能システムにおけるデータの重要性を述べ、ゲームに勝つことに比べて、言葉を理解することがコンピュータにとって難しい理由を説明します。

2019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/>
