英語コーパス学会
第 42 回大会資料

日時：2016 年 10 月 1 日（土）– 2 日（日）
会場：成城大学
（http://www.seijo.ac.jp/access/）
〒157-8511 東京都世田谷区成城 6-1-20

英語コーパス学会 2016

英語コーパス学会 第 42 回大会 プログラム

■第 1 日目

ワークショップ【British National Corpus の利用に関わる諸問題】
会　場：成城大学 7 号館 2 階 722 教室
日　時：10 月 1 日（土）10:00–12:00（9:30 受付開始）
講　師：西村　祐一
参加費：会員無料。非会員 2,000 円（当日会員としての大会参加費二日間共通）。

日　時　　2016 年 10 月 1 日（土）
受付開始　　12:00（成城大学 7 号館 2 階）
開会式　　13:00（同　7 号館 4 階 007 教室）

司　会　石井康毅（成城大学）
投野由紀夫（東京外国語大学）

1. 会長挨拶　戸部順一（成城大学学長）
2. 開催校挨拶
3. 総会
4. 学会賞審査報告　新井洋一（中央大学）
5. 事務局からの連絡

〈研究発表第 1 セッション（7 号館 2 階 721 教室）〉　司　会　宇佐美裕子（東海大学）

研究発表 1　14:00–14:30
コンピュータ環境のない英語教室における DDL のための教材開発：ハンズオン DDL と紙ベース DDL の指導実践に基づいて
若松弘子（筑波大学大学院生）・中條清美（日本大学）

研究発表 2　14:35–15:05
教育用例文を携帯端末で利用する WebSCoRE の開発とそのユーザビリティ
濱田　彰（日本大学）・Laurence Anthony（早稲田大学）・中條清美（日本大学）

研究発表 3　15:10–15:40
CEFR レベルに基づいた英単語の変換：英文難易度の最適化を目指して
内田　諭（九州大学）・高田祥平（大阪大学大学院生）・水嶋海都（大阪大学大学院生）・荒瀬由紀（大阪大学）

〈研究発表第 2 セッション（7 号館 2 階 722 教室）〉　司　会　鎌倉義士（愛知大学）

研究発表 1　14:00–14:30
コーパスを活用した get 受動態の考察
奥西嘉一（神戸学院大学非常勤講師）

研究発表 2　14:35–15:05
補文標識 for が不定詞補文の前に出現する場合の意味的特徴
西原俊明（長崎大学）

研究発表 3　15:10–15:40
共起語に見る“luxury”に込められた期待：WWW テキスト例に
近藤雪絵（立命館大学）

〈休　憩　15:40–16:00〉

シンポジウム　16:00–18:00（7 号館 4 階 007 教室）

《コーパスアノテーション（タグ付け）の功績と課題》　司　会　後藤一章（摂南大学）

- 学習者コーパスのアノテーション：「誤り」とその向こう側　講　師　和泉絵美（同志社大学）
- タグ無しコーパスとタグ付きコーパスからのコロケーション抽出　講　師　後藤一章（摂南大学）
- 語用論研究におけるアノテーション利用の現状　講　師　椎名美智（法政大学）
- 修辞項目のアノテーションを活用したテクスト分析　講　師　田畑智司（大阪大学）

《懇親会　時間：18:15–20:15　場所：7 号館地下 SEIJO LOUNGE　会費：5,000 円》

英語コーパス学会 第 42 回大会 プログラム

■第 2 日目
日　時　　2016 年 10 月 2 日（日）
受付開始　　9:10（7 号館 2 階）

〈研究発表第 3 セッション（7 号館 2 階 721 教室）〉　　司　会　山下美朋（立命館大学）
研究発表 1　　9:30–10:00
日本と米国の医学論文における論理展開の構成要素にみられる言語的特徴
―コーパスを利用した国際コミュニケーションのための学術英文の検討　　浅野元子（大阪大学大学院生）

研究発表 2　　10:05–10:35
ムーブ分析と定形表現の記述を融合する方法論の提案
―英語医学論文の導入部を例に―　　石井達也（広島大学大学院生）

研究発表 3　　10:40–11:10
学術論文のイントロダクションにおけるブースターの検証　　中谷安男（法政大学）

〈研究発表第 4 セッション（7 号館 2 階 722 教室）〉　　司　会　冬野美晴（九州大学）
研究発表 1　　9:30–10:00
前置詞句の表現分布：　　佐野　洋（東京外国語大学）・
―モノの存在形状からみた in, on, at の使用実態―　　Laurence Newbery-Payton（東京外国語大学大学院生）

研究発表 2　　10:05–10:35
上級英語学習者コーパスにみられる in/on/at/of の誤用と　　望月圭子（東京外国語大学）・
日本語の“無界性”　　Laurence Newbery-Payton（東京外国語大学大学院生）

研究発表 3　　10:40–11:10
アメリカ大統領選挙候補者の特徴　　杉山真央（大阪大学大学院生）・
―語彙使用に見る候補者のキャリア―　　木山直毅（和歌山大学非常勤講師）

研究発表 4　　11:15–11:45
TED Talk における使用語彙分析の試み　　杉森直樹（立命館大学）

〈休　憩　11:45–12:45〉

〈研究発表第 5 セッション（7 号館 2 階 721 教室）〉　　司　会　西尾美由紀（近畿大学）
研究発表 1　　12:45–13:15
Agatha Christie 作品の計量文体分析　　土村 成美（大阪大学大学院生）

研究発表 2　　13:20–13:50
Alice Bradley Sheldon 作品群の通時的著者内変化と作品の年代推定　　木村美紀（明治大学大学院生）

〈研究発表第 6 セッション（7 号館 2 階 722 教室）〉　　司　会　加野まきみ（京都産業大学）
研究発表 1　　12:45–13:15
英和辞典の記述とコーパスの活用　　田畑圭介（神戸親和女子大学）

研究発表 2　　13:20–13:50
The English Dialect Dictionary の原資料としての民俗学的情報の検討：
特にマザーグースに注目して　　谷　明信（兵庫教育大学）

講演　　14:00–15:20（7 号館 4 階 007 教室）
《New directions in corpus linguistics: Utilizing the rich annotations found in social media data》
司　会　投野由紀夫（東京外国語大学）
講　師　Laurence Anthony（早稲田大学）

閉会式　　15:20（7 号館 4 階 007 教室）
閉会の辞　　井上永幸（広島大学）

■10 月 1 日（土）
【ワークショップ】

British National Corpus の利用に関わる諸問題

西村祐一（システムエンジニア）

BNC (XML Edition)をもとにして検索ソフトを自作した経験から，BNC の Reference Guide (www.natcorp.ox.ac.uk/docs/URG/) の記述と実際の BNC のデータの間に多くの齟齬があることが判明した。BNC を研究利用するには，この点に十分注意する必要がある。

具体的には，（1）セグメント（s-unit）に付与されている番号に，同一ファイル内で重複する例があること，（2）セグメント（s-unit）および語（w-unit）に付与されたタグに構成上，あり得ないはずのものが含まれていること，（3）空白文字列のみからなる語を含む w-unit が多数あること，（4）hw (headword)が空白である w-unit が多数あること，などの問題である。

上記の（1）～（4）に適切に対処しないと，BNC-XML を直接，操作して検索する場合あるいは BNC-XML に対応した検索ソフトを作成する場合に，正確な結果が得られない可能性があることを説明する。

ワークショップは講義中心に進め, BNC-XML の検証作業は講師が操作して画面にその結果を示す。（参加者には，PC を持参して頂く必要はない。）可能な限り，技術的な予備知識がなくとも理解できるように説明する。時間に余裕がある方は，BNC の Reference Guide の少なくとも 1.3 節および 2 節に目を通して参加されたい。

今回のワークショップは BNC に特化して話をするが，コーパスが実際にどのように構成されているかを熟知することの重要性は，BNC に限らずどのコーパスにも当てはまることであり，その理解が言語研究のためのコーパス利用にとって不可欠であることを述べる。

■10 月 1 日（土）
【研究発表第 1 セッション】
【研究発表 1】

コンピュータ環境のない英語教室における DDL のための教材開発：ハンズオン DDL と紙ベース DDL の指導実践に基づいて

若松弘子（筑波大学大学院生）・中條清美（日本大学）

1. 背景・目的

本研究の目的は，コンピュータ環境のない教室においてデータ駆動型学習（data-driven learning : DDL）を実施するための教材を開発することである。DDL は，ことばの意味や文法規則をコーパスから「発見」する学習手法であり，アクティブ・ラーニングの一形態としても注目されている（赤野, 2016）。DDL の実践は，学習者自身が検索語を入力してコンコーダンスラインを調べるためにコンピュータ教室で行われることが多い。しかし，十分なコンピュータ環境を用意できなかったり，分かりやすい検索結果が表示されず学習者が混乱したりするなど，学習者がコーパスを直接検索するハンズオン DDL を授業へ導入することは必ずしも容易ではない。検索結果が印刷された紙ベースの DDL であれば，学習者に提示したい文をコーパスから取捨選択しレベルに合わせた編集も可能である（Aston, 2001）。作成したプリント教材は他の教師も使用できるなどの利便性も高い（Boulton, 2009）。本発表では，ハンズオン DDL の長所・短所に関するフィードバックの分析に基づいた，紙ベース DDL の教材開発過程とその実践報告を行う。

2. 方法

ハンズオン DDL については，2 名の教師が大学一般英語の初級クラスに対し，Web で無償公開されている SCoRE を用いて文法指導を行った。教師達は授業の振り返りを記録し，学習者の評価・感想を質問紙と自由筆記で収集した。紙ベース DDL については，1 人の教師が

SCoRE 例文に依拠したプリント教材を作成し，それを用いて文法指導を行い，教師の振り返りと学習者の評価・感想を収集した。それらを分析し，DDL 教材開発に要する時間・労力・問題点，および実践時の問題点，改善案等を洗い出した。

3. 結果

ハンズオン DDL に関する学習者側の評価では, SCoRE のコンコーダンサーが直感的に使えたため，集中して発見学習ができたなどの好意的な意見が多かった。教師側からは，検索結果を見越して，文法規則の発見に至るまでの仮説形成・検証に適した「考えさせる」タスクの作成に労力を要したという意見が出た。また，授業時に検索に伴うトラブルに直面したという指摘があった。

効率的に発見学習が可能であるという点で紙ベース DDL も学習者に好意的に受け止められた。教師側からは，ハンズオン DDL と同様に「考えさせる」タスクを作ることに加え，発見学習に適した十分な量の例文を提示する必要性も指摘された。また，教師の判断で提示したい例文を選択できる点は紙ベース DDL の長所としても受け止められた。一方の教師が作成したプリント教材をもう一方が補習用教材に転用できるという利点もあった。紙ベースとハンズオンの DDL 教材はウェブで公開されており（http://www.score-corpus.org/ の教材バンク），多くの教員とメリットを共有できることが期待される。

【研究発表 2】

教育用例文を携帯端末で利用する WebSCoRE の開発とそのユーザビリティ

濱田　彰（日本大学）・Laurence Anthony（早稲田大学）・中條　清美（日本大学）

本研究の目的は，コーパスを利用した data-driven learning (DDL) の普及に向けて，英語と日本語が併記される教育用例文コーパスを携帯端末で利用できる WebSCoRE を無償公開し，そのユーザビリティと教育効果を報告することである。DDL は，複数の事例を観察することにより，学習者が自らことばの意味や文法の規則を帰納的に発見する学習スタイルであり，アクティブ・ラーニング型の指導法のひとつとなる。先行研究では，DDL により語彙力や文法知識が向上するという教育効果が実証されている (Cobb & Boulton, 2015)。一方，授業に取り入れるには，適切なレベルのコーパスと使いやすいツールという課題に加えて，コンピュータ環境の不備という実際の教育環境の問題がある。

上記の課題を解決するために，英語初級学習者向けに開発された教育用例文コーパス SCoRE を，Web パラレルコーパス検索エンジン AntWebConc-Parallel に搭載した WebSCoRE を無償公開した。同時に，パラレルコーパスをツールに搭載する際に問題となる，日本語コーパスの分かち書きを可能にする SegmentAnt も公開した。WebSCoRE は古いパソコン教室や，Wi-Fi が強力でない教室に対応するシンプルなツールであり，携帯端末上で容易に動作する。WebSCoRE のユーザビリティを検証するため，次の研究課題に取り組んだ。

(1) コンピュータ支援英語学習に対するレディネスは WebSCoRE のユーザビリティ評価にどのように関わるか。
(2) 課題英作文における誤りの減少という観点で WebSCoRE はどれぐらい使えるか。

理系の大学に通う日本人大学生を対象に調査を行った。調査は，携帯端末を英語学習に利用できるかのレディネスを問うための質問紙，協同学習形態での課題英作文，WebSCoRE を使った課題英作文，およびアプリケーションのユーザビリティの評価で構成された。レディネスの測定には川口・草薙 (2015) の質問紙を使い，ユーザビリティに関わる評定値とどのように関わるのかを構造方程式モデリングで検証した。また，WebSCoRE が DDL 教材として使えるかどうかを，学生が感じるユーザビリティと課題英作文に見られた文法・語法の誤りの関係から統計的に検討した。

携帯端末で WebSCoRE が使える上，課題英作文を正確に成し遂げるかは，コンピュータへの慣れ親しみと，英語学習の姿勢と関連している。一方，コンコーダンスの操作の戸惑い，検索結果の読み取りなどの困難もあったことから，より高いユーザビリティを目指すことが今後の課題となる。

【研究発表 3】

CEFR レベルに基づいた英単語の変換：英文難易度の最適化を目指して

内田　諭（九州大学）・高田祥平（大阪大学大学院生）・
水嶋海都（大阪大学大学院生）・荒瀬由紀（大阪大学）

本研究の目的は，学習者に提示する英文の難易度を最適化するための手法を提案し，レベル自動調整システムの構築を目指すことである。特に学習者にとって理解の難しい単語（難単語）に焦点を当て，文脈に沿った最適な類義語を提示することを目標とする。語彙レベルの判定には，世界的な言語能力の評価指標である Common European Framework of Reference for Languages (CEFR)に準拠する。これにより，様々なバックグラウンドを持つ学習者に対して標準化した難易度を提示できる。

英文に出現する難単語は，学習者の理解の妨げとなる。Laufer (1989)によれば，適正な英文の理解を得るためには 95%以上の単語が既知であることが望ましいと指摘している。また，Krashen (1985)は，インプット仮説を提唱し，言語学習がもっとも効率的に行われるのは，学習者自身のレベルよりも僅かに高い英文（インプット）を理解したときであると指摘している。このような観点は教員が学習者に提示する英文を選ぶ際に重要であり，英文が難単語を多く含んでいれば，注釈をつけたり，難単語を置き換えたりという処置が必要となる。しかしながら，これらの手続きは極めて経験的であると同時に多くの時間と労力を要する。

本研究は，目的レベルに合致した英単語の選定作業を支援するためのシステムの構築を目指すもので，共起スコアをベースにした客観的な言い換えの手法を提案する。文中の単語の言い換えの研究は自然言語処理の分野で盛んに行われているが（Biran et al. 2011, Horn et al. 2014 など），教育現場のニーズを反映したものは少なく，現場レベルの教員や教材作成者が手軽に利用できるものはほとんど存在しない。本研究の提案手法は，言語教育のニーズに沿って真に実用的なシステムの構築を目指し，CEFR レベルに基づいて英単語の難易度を調整するという枠組みのもと，統計的手法を用いて言い換えの候補を絞り込む。具体的には，次のような手順を踏む。

(1) TARGET の選定：CEFR レベル B2 以上（C2,C1,B2）のものを言い換えの対象とする。CEFR レベルは主に「CEFR-J Wordlist Version 1」（2013: 東京外国語大学投野由紀夫研究室）に基づく

◆One of the most common misconceptions (B2: TARGET) about …

(2) 類義語の抽出：電子データで利用可能な類語辞書（Roget's 21st Century Thesaurus, Third Edition）から Target の類義語を抽出し，CEFR レベルで絞り込む(SYNONYMS)

◆misconception→fallacy (NA), misinterpretation (C2), fault (A2: SYNONYMS), mistake (A2: SYNONYMS) etc.

(3) HEAD の探索：Target の前後の修飾関係を機械的に探索し，HEAD を特定する

◆One of the most common (adj: HEAD) misconceptions (noun) about

(4) 共起スコアの算出：HEAD と SYNONYMS の共起スコア（Corpus of Contemporary American English (COCA)のデータを使用）を算出する

◆common + fault = 0.21, common + mistake = 3.32

この結果，misconception は SYNONYMS のうち mistake ともっとも共起スコアが高く，置き換え可能と判定される。本発表では，これらの置き換えについてネイティブスピーカーによ

る判定結果を提示し，その精度をさらに高めるための手法について議論する。また，ウェブシステムとしての実装の経過についても報告する。

■　10 月 1 日（土）
【研究発表第 2 セッション】
【研究発表 1】

コーパスを活用した get 受動態の考察

奥西嘉一（神戸学院大学非常勤講師）

get 受動態に関して，以下のよう２つのリサーチクェスチョンを設定し，BNC を使って調査・研究を行った。

1. get 受動態は，純粋に中立的な状況を表す受動態より，むしろ主語がさしているもの，もしくはそれと関係している誰かに対して利害関係を表現する受動態において独占的に生起してくる(Huddlestone & Pullum, 2002)とされているが，実際のところ本当にそうであるのか。まず BNC の Spoken Texts からランダムオーダーで get 受動態と be 受動態の文を２５０取り出し，get 受動態の文の中の動詞の頻度を調べ，３以上の動詞をリストアップした。そしてそれらの動詞について１００万語当たりの hit 数が be 受動態より多い動詞を選び, それらの動詞を含む全ての文について調べた。
2. get 受動態は，よりインフォーマルなコンテクストの中で使われ，文語英語より口語英語においてより一般的である(Huddlestone & Pullum, 2002)と言われているが，現代英語において get 受動態は本当に文語英語より口語英語において使用頻度が高いのか，高いとすればどれほど高いのか。BNC の Spoken Texts と Written Texts のそれぞれにおける get 受動態の hit 数と１００万語当たりの get 受動態の hit 数を調べ比較した。

上記２つのリサーチクェスチョンを解決するために，使用する BNC における検索式を be 受動態は｛be/V｝_VVN ，get 受動態は｛get/V｝_VVN とした。

１～２のリサーチクェスチョンの結果を以下に記す。

1. get 受動態は「主語がさしているもの，もしくはそれと関係している誰かに対し利害関係を表現する受動態において独占的に生起してくる。」のではなくむしろ「純粋に中立的な状況を表す受動態」においてよく使われる。

参考例文：Do you know if she got married?（結婚したかどうかを問題にしているだけである。）

2. get 受動態は hit 数では口語英語より文語英語の方が多い（約２倍）が，１００万語当たりの hit 数（頻度）は文語英語より口語英語の方がはるかに多い（約４.５倍）。

【研究発表 2】

補文標識 for が出現する場合の意味的特徴

西原俊明（長崎大学）

英語の動詞には，(1)のように不定詞補文をとり，for-句の生起を許す動詞が存在する。

(1) I want very much for you to stay right here.　(稲田(1989: 54)

この形式をとる動詞に関して，稲田(1989)が可能な動詞のリストを示しているが，リストにない他の動詞，形容詞もこの形式が可能である。また，Bošković (1992)は，Bresnan (1972)の分析にふれ，意味的な役割により for-句が生じると述べるにとどまり，意味的特徴については詳しく述べられていない。さらに，統語論研究の多くは，(1)のように修飾語句が動詞と補文主語との間に介在する場合，for が生起しやすいことが指摘されている。しかしながら，(2)に示すように隣接性とは関係なく for-句は生じることができる。

(2) a. For years, fans yelled for the Giants to use and develop more young players.　(GloWbE)
　　b. We saw her in the tall glass, and I motioned for her to stay.　(COCA)

c. She squeezed Julie's hand and motioned for her to follow down the stairs that led to the ground floor. (BNC)

本発表では，COCA，BNC やニュース記事に見られる表現をもとに for-句を伴う不定詞補文の使用域を明らかにする。また，for を伴わない場合とでは意味的に異なる特徴が存在することを示し，その意味的特徴を明らかにする。動詞及び形容詞を二つの類に分け，指示を出す signal 型(signal, motion, gesture, scream, shout, yell, clamor, wave)と感情伝達型 root, push, craving, desperate, keen に分けて for-句を伴う場合の意味的特徴に関して考察を行う。

【研究発表 3】

共起語に見る *luxury* に込められた期待：WWW テキスト例に

近藤雪絵（立命館大学）

近年 *luxury* は，その語が持つファッショナブルな響きにより，広告等で乱用される傾向にある（Kapferer 2012）。本研究はある語が乱用されるのは，その語が効果的に我々の期待に働きかけるという想定があると考え，*luxury* が持つ性質を，その修飾する名詞と等位接続詞により並列共起する名詞から探求し，我々が luxury を何に見出し，何を得ようと期待するのかを明らかにしようとした。分析対象は近年の WWW テキストとし，大規模ウェブクローリングコーパスである enTenTen 及び ukWaC を用いた。分析には Sketch Engine を用いた。

luxury が修飾する名詞には主として「休暇」に用いられるもの（例：*hotel*, *villa*,）が抽出された。また *luxury* が「日常」に用いられるもの（例：*car*, *watch*）を修飾する際，加えて特別性を示す語を伴うことが分かった（例：*luxury sport car*）。これにより，我々は luxury を，1. 日常から切り離された環境（休暇等）の中に或いは 2. 日常品に特別性（性能，ブランド，材質）を加えることで見出していることが示唆された。2.の日常品をさらに詳しく見るため *luxury item(s)/product(s) such as* X を検索したところ，このフレーズの中で *luxury* と高頻度で共起語する一方で，特別性を示す語を伴わない名詞として *perfume*, *jewelry* が抽出された。これらは luxury を象徴する物であるといえる。上述の名詞は全て滞在したり，乗り込んだり，着用したりするものであることから，我々は自身とそのものを一体化することで luxury を得ようとすると考えられる。

次に，Sketch Engine の Word Sketch 機能を用い，*luxury* と並列共起（*luxury and/or/,* X 及び X *and/or/, luxury*）する語を類義語で分類したところ，大きく *comfort*, *elegance*, *opulence* のクラスターに分類された。並列に共起する場合，使用する接続詞や語順によって意味が変わるの可能性は否めないものの，大きくみてこれらの語は luxury なものが併せ持つと期待される性質を意味すると考えられる。これら 3 つのクラスターに分類された語から，luxury は 1. 環境を快適で安心できるものにし（例：*comfort*, *convenience*），2. 環境や自身を洗練された人物にし(例：*elegance*, *sophistication*)，3. 豪華さや豊かさを感じさせる(例：*grandeur*, *splendor*)性質を併せ持つことが推察された。

本研究は *luxury* の共起語から，その語が持つ性質を探り，我々の期待，すなわち luxury な製品を身につけたりサービスを受けたりすることで我々は何を得ようと期待するのかを関連付けた。今回は WWW テキストのみを対象としたため，広告主，消費者等，視点を一つに絞るには至らなかったが，luxury な製品・サービスの広告やレビューを対象としてこのように乱用される語を分析することにより，そこに込められた期待，さらには欲求や意図を考察することに応用できると考えられる。

■10 月 1 日（土）
【シンポジウム】

コーパスアノテーション（タグ付け）の功績と課題

司会　後藤一章（摂南大学）

コーパス言語学研究において最も重要な課題の一つは検索である。いかなる分析を行うにせよ，任意の言語項目の検索処理はすべての起点となる。多くの場合，これには文字列パタンマッチングが用いられ，単語や単語列の検索処理，場合によってはその頻度計測も同時に行われる。しかし，単純な文字列検索は低コストで汎用性が高いという利点があるものの，より抽象度の高い統語情報や修辞情報，まして，高度な主観的判断を必要とする言語使用の誤り情報や語用論情報などを過不足なく検索することは難しい。こうした情報の検索にはアノテーション，すなわちテキストや語句へのメタ言語情報の付与が必要となる。近年のコーパス研究の発展とともにコーパスの用途も多様化し，様々な情報検索ニーズを満たすためにもアノテーションの重要性はますます高まっている。そこで本シンポジウムでは，これまで各自の研究目的に沿ってアノテーションを活用してきた 4 名の発表者が，その功罪に触れながら，各分野のアノテーションの現状について紹介する。

学習者コーパスのアノテーション:「誤り」とその向こう側

和泉絵美（同志社大学）

本発表では，誤り情報付与に代表される学習者コーパスのアノテーションに関して，近年の研究動向および今後期待される展開を概観する。誤り情報は，学習者コーパス研究が行われ始めた 1990 年代より，学習者コーパス特有のアノテーション対象として中心的に扱われてきたが，学習者言語に見られる現象を母語話者言語との比較に基づいて逸脱として扱うことは，ありのままの学習者言語の姿を捉え損ねる原因となるという論点で批判されることも多い。本発表では，そのような批判に含まれる一定の真理が，誤り情報付与の具体的な改善にどのように貢献してきたか検証する。また，近年 CEFR に代表される Can-do ベースの言語教育が隆盛を見せ始めていることに伴い，学習者コーパスにも新たな役割が課されつつある。「できないこと」に焦点を当てる誤り情報付与に対して，学習者がその言語を使って「できること」を見出すための新たなアノテーションに関する展望についても述べる。

タグ無しコーパスとタグ付きコーパスからのコロケーション抽出

後藤一章（摂南大学）

タグ無しコーパスからコロケーションを抽出する際，しばしば共起スパンを用いた方法が採用される。当該手法は特定の確率的言語モデルに依存せず，探索的なコロケーションの検出などに有効だが，受動態や埋め込みなどによって生じる不規則な語順変化の影響を受けやすい欠点がある。一方，アノテーションが施されたコーパス，特に統語情報が付与されたタグ付きコーパスからコロケーションを抽出する場合，句構造や単語間の依存関係などが既に明らかになっているため，語順の影響を受けず，信頼性の高い安定したコロケーション抽出が可能になると考えられる。ただし，タグ無しコーパスとタグ付きコーパスにおいて，抽出されるコロケーションにどのような違いが現れるかは実際にはあまり明らかにされていない。また，統語タグを付与する統語解析技術は高度な数理モデルに基づいており，解析結果

のフォーマットも複雑なため，コーパス言語学研究においてその利用は十分に進んでいるとは言い難い。そこで本発表では，タグ無しコーパスとタグ付きコーパスからのコロケーション抽出を比較しながら，統語解析の使用メリットについても検討したい。

語用論研究におけるアノテーション利用の現状

椎名美智（法政大学）

本発表では，ランカスター大学で共同製作した社会・語用論コーパスのアノテーションについて概説し，そのメリットとデメリットについて論じる。またそれを基礎にして，独自の研究テーマに合わせて考案したアノテーションを付けたコーパスとその研究例を紹介しながら，語用論研究におけるアノテーション利用の現状，その利点と問題点について考察する。

発表者の作った Vocative-focussed Sociopragmatic Corpus は，「テキスト解釈」という読み手の主観と解釈を含むアナログ情報をアノテーションというデジタル情報へと変換する作業の結果できあがったコーパスである。研究への応用の幅は広いが，作成過程における問題は少なくない。それをふまえて，今後の研究の展望と方向性，そして共同研究の可能性を探っていきたい。

修辞項目のアノテーションを活用したテクスト分析

田畑智司（大阪大学）

テクスト計量研究の多くは単語の生起頻度を変数として用いる‘a bag of words’（語彙頻度ベクトル）モデル(Juola, 2006)に依拠している。語彙頻度ベクトルに基づく分析は，単語の形式についての定義さえ明確であれば，低コストで計量でき，かつ，高い精度でテクストを分類・識別できるため，アノテーションなしで行われることが多い。利用されるアノテーションはせいぜい品詞タグ(POS タグ)にとどまっているものがほとんどである。より抽象的な，言語項目の表現機能・修辞的効果は，形式に基づく一義的な定義が困難であり，テクストにそうした情報を埋め込むことには大きなチャレンジが伴うため，その実装は容易ではない。

そのような状況下，言語表現のプライミング効果を基準にして，修辞項目の階層的カテゴリーを提示する Kaufer et al. (2004)の研究は，単語レベルだけでは計量困難なテクストの特徴を量的に記述する手がかりを提供しようという野心的な試みであり注目に値する。本発表では，Kaufer, et al. (2004)の Language Action Types (LATs)のアノテーションを基に，テクストの量的特徴付けを行う。具体的には，どのような修辞カテゴリーがテクストの分類・識別に寄与するかを視覚的に示すだけではなく，特徴とされる LATs がテクスト中でどのような機能を担っているのか，質的読みを組み合わせ，local textual functions に着目した議論を展開したい。

■10 月 2 日（日）
【研究発表第 3 セッション】
【研究発表 1】

日本と米国の医学論文における論理展開の構成要素にみられる言語的特徴 —コーパスを利用した国際コミュニケーションのための学術英文の検討

浅野元子（大阪大学大学院生）

本発表の目的は，日本と米国の同一専門分野の英文医学誌の論文テクストを計量し，専門英語教育の観点から考察部分のムーブを分析して，知見を各国の読み手に伝える英語における特徴を明らかにすることである。日本の循環器学誌 *Circulation Journal* (*CircJ*) と世界的に著名な米国の循環器学誌 *Circulation* (*Circ*) の研究論文各 10 報をコーパス化した。

国際コミュニケーションで使用される英語が国際英語と定義され (Smith, 1976)，英語使用者間での相互理解が可能な統一性と多様性を有する英語変種から成ることが想定されているが (日野, 2008)，近年，学術論文が英語で発信される頻度が増しており，若手研究者や大学院生が英語で論文を書く機会が増えている (Swales & Feak, 2012)。日本では文部科学省が臨床医学論文の出版を医学教育での課題の 1 つに挙げている (文部科学省, 2010)。日本の使用者のための国際英語モデルの構築が説かれるが (日野, 2013)，研究論文は国際英語としては Discipline-specific な英語の変種とされ (Clancy, 2010)，米国の *Circ* は，科学での文化 "scientific culture" (Swales, 1990: 65) の中でも循環器学分野の文化による多様性を有する国際英語の例と考えられ，日米の循環器学論文を比較し当該ディスコース・コミュニティのメンバーに受容される英語の特徴を検討することは教育的に意義があると考えられた。

日米の論文テクスト全体では，クラスター分析を組み合わせたコンセンサス・ツリーで判別され，差異が示唆された。日本の論文での高頻度語を田畑 (2016) に従い Mann-Whitney の U 検定で検討しコンコーダンス・ライン上で精査すると "evaluate" や "the present study" など日本の教科書にみられる表現が散見された。

考察部分は，発表者と 2 名の医学専門家が Nwogu (1997) などを参考にムーブを分析し，テクストを検討すると，2 誌ともにムーブの下位要素であるステップによる言語的特徴が示唆され，結果の叙述に "found" などの動詞，結果の解釈や限界の叙述に "might" などの法助動詞が高頻度に認められた。各論文でのステップの出現順序を図表化すると論理の流れにパターンが認められると考えられた。

日本と米国の論文は，テクスト全体では異なる言語的特徴を有する可能性が示唆されたが考察部分ではステップによる類似性が示唆された。本研究での知見は教育的な意味があると考えられ，今後さらに検討を続けたい。

【研究発表 2】

ムーブ分析と定形表現の記述を融合する方法論の提案 －英語医学論文の導入部を例に－

石井達也（広島大学大学院生）

English for Specific Purposes (ESP)におけるジャンル分析では，ある特定の共同体における文章構成を明らかにするためにムーブ分析が行われてきた (Coffin, 2001; Hyland, 2002)。ムーブ分析においては，共通の知識と，それを具現化する言語様式が共有されているはずであるという前提がある (Partington et. al., 2014)。一方でコーパス言語学では，特に新 Firth 学派において，語彙と文法は切っても切り離せない関係であるという考えと共に，新たな言語分析単位として，定型表現 (phraseology)という概念を発展させてきた (Hunston & Francis, 1999)。ある特定の共同体がムーブの流れと，それを具現化する言語様式を共有しているならば，そこで用いられている文章には，共通で用いられている定型表現があるはずである。しかしながら，コーパスデータを用いてムーブの流れを明確にする定型表現の記述の試みは少ない。そこで

本発表では，コーパスデータを用いて，ムーブの流れを明確にする定型表現を記述するための方法論と，その方法論の実践として英語医学論文の導入部を用いた具体的な結果について発表する。

ムーブの流れを明確にする定形表現を記述するために，ムーブごとにコーパスデータを集積する。その後，参照コーパスを全コーパスとし，それぞれムーブごとのキーワードを算出し，分析する。その際，頻出度が一番多い機能語の振舞いに着目し，WordSmith5 を用い，ランダムに示された 100 文をコーパス駆動型の手法で分析する。以下具体的に英語医学論文の導入部を分析実践とし，その具体的な結果の一部を記述したい。

Nwogu (1997) が提唱する医学論文のムーブの流れに基づき，2013 年及び 2014 年に発行された Impact Factor の高い 4 つの国際雑誌から医学論文 395 部の導入部を, 3 つのムーブに分割し，コーパスデータを構築した。3 つのムーブはそれぞれ，背景知識の導入(ムーブ 1)，先行研究とその問題点の指摘 (ムーブ 2)，研究の目的とその主たる研究手法の提示(ムーブ 3)であった。その後上記で示した手法を用いた。その結果，各ムーブに関連する定型表現をそれぞれ 6 つずつ記述できた。ここでは紙面の都合上，ムーブ 1 における一部の結果のみ記載しておく。ムーブ 1 のキーワードは be 動詞 *is* であった。be 動詞 *is* の前後の関係を見ると，be 動詞 *is* の前では論文のトピックを主語とし，be 動詞 *is* の後では the most common cause of や the leading cause of といった表現と共起していた。これらの定型表現はムーブ 1 の機能である背景知識の導入のために用いられていた。本発表では，これらの定型表現を含め，医学論文の導入部におけるムーブの流れを明確にする 18 の定型表現を提示する。

【研究発表 3】

学術論文のイントロダクションにおけるブースターの検証

中谷安男（法政大学）

1.はじめに

Swales (1990, 2004)などで主張されているように，学術論文のイントロダクションにおいては，ムーヴ (Move) を確立し読者を引きつけることが重要である。特に，研究が 十分新規的な内容で興味深いこと，また研究課題の重要さについて端的に述べることが必須の要素と認識されている（例 Vassileva, 2001）。このためには，書き手の主張や断定を強めるブースター(Booster)の活用が示唆されている。しかし，先行研究においては，具体的にどのようなブースターがいかなる目的でイントロダクションの箇所で使われるのか十分議論されているとは言えない。本論ではこの点に注目し，実際の学術論文を収集した英文コーパスを作成し，イントロダクションにおけるブースターの活用を検証した。

2.検証方法

研究論文の一般的な傾向をみるため，自然科学，社会科学の経済・経営，人文科学の応用言語学から, 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 2 つずつ選んだ。*Science, Nature, International Economic Review, Journal of Management, Modern Language Journal, Language Learning* の 6 誌の 2006 年より 2011 年に掲載された研究論文の中から，第一著者が英語ネイティブと思われる 17 本をそれぞれ選んだ。これらを電子ジャーナルからダウンロードしテキストファイルに変換した。この合計 102 本の論文による総語数 105 万語の学術論文コーパスを作成した。

この中の Introduction として明記している章，または明記されていない場合は，それと同等の最初の章の総計 79,876 語を抜き出した。このイントロダクションのコーパスを，学術論文コーパスの他の部分を参照コーパスとして，Wordsmith 6.0 を活用し特徴語を抽出しクラスター表現を確認した。また Hyland（2005）のブースターのリストを活用し，イントロダクションにおけるこれらの表現の頻度と，参照コーパスにおける頻度と比較した。

3.結果

分析の結果，イントロダクションでは特定のブースターのクラスター表現が頻繁に使われていた。これらを Allison (1995), Hyland(2000, 2005)等の先行研究に基づき手動で分類すると，以下のような 5 つのグループの表現に集約することが可能となった。

①範囲の広さ：広く認められている研究領域という主張　例 It is widely believed that
②数の多さ：研究の数が多い重要な課題　例 A number of studies have been conducted
③期間の長さ：長い間取り組まれている研究　例 research has long been recognized that
④新規性：最新の研究領域だと重要性を強調　例 Recent research trends towards
⑤ポジティブさ：ポジティブな表現で内容を肯定　例 There is compelling evidence that

本発表では，これらの分類に該当する事例を検証していく。

■10 月 2 日（日）
【研究発表第 4 セッション】
【研究発表 1】

前置詞句の表現分布　－モノの存在形状からみた in, on, at の使用実態－

佐野洋（東京外国語大学）・Laurence Newbery-Payton（東京外国語大学学部生）

英語の前置詞句表現は，日本語母語話者にとって学習困難点の一つである。日本語と英語で，コトを表す時空間の認識次元が異なることが要因の一つである。事実，関係性を表すモノと出来事をつなぐ日本語の助詞と，存在を表すモノの出来事における役割を示す英語の前置詞の種類数には著しい差がある。空間の表現視点からは一対多の関係になるから，一つの成分（日本語の助詞付き表現）の対応先が，複数の句（英語の前置詞句）になる。表現を仕分ける条件の習得の厄介さが学習困難を生み出していると考えられる。

本発表では，BNC を用いて，前置詞句内の名詞句表現の分布を調査したので報告する。前置詞が，名詞の意味の概念化に，どう影響するのかを明らかにすることを目的としている。

時空間内の存在物の外形特徴は，抽象化や一般化した性質として区分できる（表 1)。表 1 を基に形状属性を考え，前置詞は，名詞が表すモノの形状に関連すると仮定する。三次元的な世界ではモノの動きもあると考えると表 2 のように前置詞と空間次元を対応させることができ，対応する名詞の意味を空間次元の性質から限定したり制限したりする。

表 1

一般（抽象）化した存在形状	形状属性や外形特徴
立体（三次元）	立体（非幾何学的立体と幾何学的立体），立体内と外
平面（二次元）	平面（非幾何学的平面と幾何学的平面），平面内（領域内）と外，立体表面
線（一次元）	線（非幾何学的直線と幾何学的直線），線の上，線全体
点（ゼロ次元）	点，点の近傍，点の周囲

表 2

前置詞	存在形状	モノの存在の意味	モノの機能，役割の意味
in	三次元，二次元	内部構造，分野範囲，時間（幅，区間）	様態（非意志的）
on	二次元，一次元	具体物，接触，時間（幅，区間全体）	時間，活動（意志的）
at	点（ゼロ次元）	点，時間（時刻）	活動（非意志的）

表 2 に従う（英語の）名詞の意味分布があるのか，それら分布実態の特徴を，BNC を用い

て名詞に後続する前置詞句から調査した。調査には，NLTK（Natural Language Toolkit, python）を利用した。

各前置詞（in，on，at）に後続する名詞句は，(1) 無冠詞名詞（複数形を含む），(2) 不定冠詞付き名詞，(3) 不定冠詞と形容詞付き名詞である。

およそ 4,280 万語の調査結果では，抽出数，“in Noun”/47924，“in a Noun”/8651，“in a ADJ Noun”/5593，“on Noun”/6338，“on a Noun”/17942，“on a ADJ Noun”/3292，“at Noun”/8906，“at a Noun”/2179，“at a ADJ Noun”/1329 である。(1) のタイプの頻度が多い。各前置詞で分布特徴が類似していた。どの前置詞句も無冠詞の用法が多く，慣用的で定型表現的な表現が多用されていることを表している（と考えられる）。

前置詞句単位の語彙密度（表現のバラツキ）から，表現の多様性は不定冠詞と形容詞が前接する名詞句が多いことが分かる。無冠詞の用法は絶対数が多いが，表現の多様性は少ない。(3)のタイプの種類の多さから形容詞を用いて存在の意味を表していると考えられる。

表 2 に基づき，前置詞と名詞の意味分類の共起制限の特徴について議論する。

【研究発表 2】

上級英語学習者コーパスにみられる in/on/at/of の誤用と日本語の“無界性”

望月圭子（東京外国語大学）・Laurence Newbery-Payton（東京外国語大学学部生）

本発表の目的は，日本語を母語とする上級英語学習者コーパスにみられる「前置詞の誤用・非用」の計量的分析を通して，前置詞の誤用・非用の実態を観察し，それがどのような日本語の特性と関わっているのかについて考察する点にある。

東京外国語大学英語専攻学生による上級学習者コーパスに基づく「オンライン英作文学習者コーパス・誤用辞典 Online Dictionary of Misused English」（http://sano.tufs.ac.jp/lcshare/で公開中）において，前置詞の誤用・非用・正用数とその割合は，以下のとおりである。

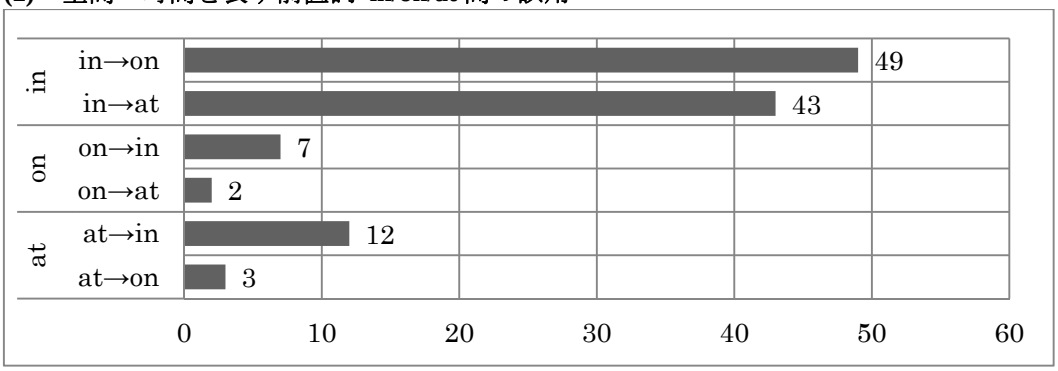
(1) ODME における前置詞の誤用・非用・正用

前置詞	総頻度	誤用	非用	正用	正用･誤用比（%）	正用･非用比（%）
in	5419	279	110	5140	5.4280156	2.1400778
on	1408	37	199	1371	2.69876	14.514953
at	776	36	67	740	4.8648649	9.0540541
of	6607	162	82	6445	2.5135764	1.2723041
for	2312	46	73	2266	2.0300088	3.2215357
to	2255	47	50	2208	2.1286232	2.2644928
from	1195	17	19	1178	1.4431239	1.6129032

(1) から，過剰使用による誤用率が高い前置詞は，in, at であり，また非用率が顕著なのは，on, at であることがわかる。

次に，以下の(2)では，on/at とすべきところに，in を用いた誤用傾向が観察される。

(2)　空間・時間を表す前置詞 in/on/at 間の誤用



(1)及び(2)が示すように，in の誤用が最も顕著であるが，なぜ「“in”の過剰使用」が日本語母語学習者に顕著なのだろうか。その要因として，英語では，in/on/at という空間的認知が明確に存在するのに対して，日本語の場所を表す表現は，述語の項構造によって「に」と「で」の使い分けが決まっていて，明確な空間認知と無関係であるという相違が考えられる。日本語母語話者にとって，事物や事象を，空間認知の視点からとらえることはむずかしいという現象が，前置詞の誤用につながっていると推測される。

次に，(1)で誤用頻度が 4 番目に高い「“of”の過剰使用」は，内部構造の in, 移動表現の起点としての from, 未完結の for といった前置詞を使うべきところに,「“of”の過剰使用」が観察されるが,「NP_1 の NP_2」という連体修飾表現が，日本語母語話者の“of ”の過剰使用の原因になっていると推測される。学習者コーパスの対照から日本語母語話者は，中国語母語話者に比較して,“of ”の過剰使用が卓越していることも，観察される。本発表では，日本語における空間・時間的認知の“無界性”が,前置詞の習得を困難にしていることを明かにする。

【研究発表 3】

アメリカ大統領選挙候補者の特徴 —語彙使用に見る候補者のキャリア—

杉山真央（大阪大学大学院生）・木山直毅（和歌山大学非常勤）

本研究は，アメリカ大統領選挙候補者のスピーチをコーパス化し，大統領選挙演説でのスタイルが候補者のキャリアが影響していることを提案する。

これまで政治家の表現に含意される意味（レトリック表現など）について多くの研究が行われてきた（e.g., Fairclough，2010; Van Dijk，2008; Scott，2008）。このような研究から，相対する政党と比較することで自らの視点の重要性を強調する傾向があることが明らかになっている（Beads，2000: pp. 24）。

以上の先行研究を参考に，本研究では 2016 年度大統領選挙において，最後まで候補として残っていたドナルド・トランプ氏，ヒラリー・クリントン氏，バーニー・サンダース氏の選挙演説において各氏がどのようなスピーチスタイルを採用しているのかを分析する。

本研究は各候補者から 5 本ずつの大統領選挙演説を使用した。それらの高頻度語 150 語を対応分析すると，共和党と民主党のクラスターに分けられる。トランプ氏は *don't, we're* などの縮約（=（1））, *very* などの強意副詞が特徴として見られた（=（2））。一人称では *I* の使用が特徴として見られる。

（1） I don't want people to go around thinking that I have a problem. (February 24, Trump)
（2） I think you're going to find it very informative and very, very interesting. (June 7, Trump)

一方，サンダース氏には *energy* や *economics* といった具体的な政策に関する語彙が，クリントン氏には *need* や *more* など，米国に必要な事柄，発展を示す語彙が特徴として見られた

(=（3，4))。また，両者は一人称主語に *we* を用いる傾向にある。

（3） We have a moral responsibility to work with countries throughout the world to transform our energy system away from […] sustainable energy. (February 10, Sanders)
（4） We believe we need to make America the clean energy superpower of the 21st century […]. (June 7, Clinton)

以上の観察に基づくと，各候補者の演説について，各氏のキャリアが関わっていることが示唆される。トランプ氏は，長く米国ビジネス界のリーダー的存在でであったため，口語表現，自身の考えを述べる *I* の使用は，ビジネスリーダーとしての影響であると考えられる。一方，サンダース氏とクリントン氏は，現在の米国における問題点，政策を述べている。上記に加え，彼らの演説には *we must, need to* が特徴として見られた。これは，サンダース氏やクリントン氏は，政治家としての経験が長く，近年の大統領が国民とのビジョンを共有する方略を採用しているものと考えられる（田畑，2012)。

本研究では，トランプ氏，サンダース氏，クリントン氏のスピーチスタイルが，各氏が積んできたキャリアに起因することを論じた。

【研究発表 4】

TED Talk における使用語彙分析の試み

杉森　直樹（立命館大学）

TED (Technology, Entertainment, Design)は，さまざまな分野における活動家や思想家，アーティスト，研究者等がスピーチを行うイベントとして知られているが，NHK の番組放送に加えて，最近では TED Talk を題材として使用した英語の教科書も出版されるなど，英語学習の教材としても注目されてきている。TED Talk のスピーカーは，Gallo (2014)が示しているような TED Talk 特有のスピーチのスタイルに基づいて話すことが多いと考えられるが，それらが言語的にどの様な特徴を持っているかに関するコーパスを用いた研究はまだ十分には行われていない。本研究は，TED Talk のスピーチの特徴がその使用語彙の中にどの様に反映されているかについての分析を試みたものである。

本研究では，Official TED Talk Guide Playlist 等において代表的な TED Talk として選定されているものを中心に, 56 本の transcript を TED のサイトから収集し, 総語数約 160,000 語のデータを得た。これらのデータに対して AntConc 及び AntWordProfiler を用いて使用語彙の頻度分析とレベル分析を行った。また，学術英語としての特性を検証するため，大学の講義やセミナーを集めたコーパスである British Academic Spoken English (BASE) Corpus との比較を行い，対数尤度比による特徴語の分析を行った。その結果，TED Talk においては，Donovan (2014)や Wang (2012)が述べているように，*I* や *we* 等の人称代名詞が特徴的に使用される傾向があることが本研究においても確認された。また，*when, people, story* といった TED talk において特徴的に使用されると考えられる語彙の存在が認められた。これらの分析結果は，TED Talk が public speaking の要素を持ち, その分野の専門家が一般聴衆に自分の主張を効果的に訴えかけるために storytelling の手法を用いる傾向があることを示唆している。本研究発表では，これらのデータや用例を示しながら TED Talk の持つ語彙的特徴に関する考察を述べる。

■10 月 2 日（日）
【研究発表第 5 セッション】
【研究発表 1】

Agatha Christie 作品の計量文体分析

土村成美（大阪大学大学院生）

本研究ではミステリー作家 Agatha Christie の作品の文体を, 他のミステリー作家との比較を通して明らかにすることを目的とする。比較対象としては, ジャンルによる差異を最小限にするためにミステリー作品のみを用い, Christie と交友があったとともに良きライバルでもあった Dorothy L. Sayers, ミステリーの礎を築いた一人である Arthur Conan Doylle の作品を用いる。

Sayers は, 同時代に活躍していた Christie と長編作品の初版部数が拮抗して人気を二分しており (Malling & Peters, 1998), Christie と並んでミステリーの 2 大女王と称され, Christie と共に語られることが多い。Doylle については, Christie が幼少期に姉と共に Sherlock Holmes 作品を読んでおり(Morgan, 1986), 彼女の文体に影響を及ぼしている可能性があるかもしれないため選定した。

Christie 作品の計量的な分析としては, 語彙多様性や曖昧語に関する分析を行った Lancashire & Hirst(2009)や, 語彙面と統語面からアプローチを行った Le et al.(2011)などが挙げられるが, 用いられている計量的な指標は語の使用率や Type/Token Ratio のみで, 多変量解析を行った研究は見当たらない。

本研究では

1.Christie, Sayers, Doylle の作品は統計的な手法を用いた際に, どれ程の正解率で分類を行うことが可能か

2.分類に寄与した特徴語のうち, Christie 作品における特徴語は, 他の 2 作家とどのように異なる使用がされているのか

以上の 2 つのリサーチクエスチョンに対して分析, 検討を行う。

分析データとして, Christie221 作品(5,230,256 語), Sayers55 作品(1,430,257 語), Doylle60 作品(667,901 語)を使用し, 機械学習の一種である Random Forests を用いて分類と特徴語抽出を行った。Random Forests は Breiman(2001)が提唱した分類, 回帰を行う手法である。Random Forests は元のデータからブートストラップサンプリングされたデータを用いて決定木を生成する。無作為抽出されたデータの 3 分の 2 を用いて判別モデルを作成し, 残りの 3 分の 1 を用いてモデルのテストを行う。金・村上(2007)では, テキスト分類の際に他の機械学習の手法に比べて Random Forests の分類精度が高いことが報告されている。また田畑(2012)では, Random Forests を Charles Dickens と Wilkie Collins を区別する特徴語抽出の手法として用いており, 従来の統計指標(カイ二乗値や対数尤度比)を用いた特徴語抽出に比べて, 信頼性の高い特徴語の抽出が可能であることを示している。本研究においても Random Forests を用いた作家の分類を行うと同時に, 特徴語の抽出を試みる。

分析指標には著者推定研究において有効性が示されている高頻度語を用い, 用いる語を 100 語から 1000 語まで 100 語ずつ変化させ, 分析を行った。分類精度は 92.9%~97.3%であった。最も高い平均分類精度が得られた変数が 500 語の時の RandomForests 実行結果をもとに, 他の 2 作家と比較した際の Christie の使用語彙の特徴について考察を行う。分析の結果, Christie 対 Sayers の *someone*/*somebody*, *anyone*/*anybody* のような類義語の使用の違いや, Christie 対 Doylle の動詞縮約形の使用率の違いといった語彙的特徴が明らかになった。

【研究発表 2】

Alice Bradley Sheldon 作品群の通時的著者内変化と作品の年代推定

木村美紀（明治大学大学院生）

本研究では，コーパス言語学の一分野である計量文体論の手法を用いて Alice Bradley Sheldon 作品群に関する通時的な著者内変異の検出を行う。Alice Bradley Sheldon (1915-1987: 米国) はデビューから約 10 年間，正体不明・性別不明の作家 James Tiptree, Jr.と Raccoona Sheldon として著作活動を行ってきた作家である。文芸批評上ではこの作家の文体が正体露見という出来事を契機に変化していると論じられることが多い。とりわけ, 小谷 (1999: 84)では，「SF 界の賞を総嘗めにしたものの，1976 年の正体露見以後は作風も変化し，年齢もあって執筆量は減少した」といったように， Alice Bradley Sheldon 作品群全体に対する通時的な文体変化の可能性を示唆している。

この作家の通時的な文体変化を定量的に検証するため，Alice Bradley Sheldon 作品群 72 作品を含めたコーパスを構築した。先行研究において実行の容易さと指標の有効さが指摘されている高頻度語彙を指標として採用し，執筆年・出版年の判明している 71 作品に対し，主に執筆年を基準として分析を行った。これら 71 作品を，1960 年代作品群，1970 年代初期作品群，1970 年代後期作品群，1980 年代前期作品群，1980 年代後期作品群という 5 カテゴリーに分けた。各カテゴリーのサイズは，60 年代作品群が 22 作品，70 年代初期作品群が 22 作品，70 年代後期作品群が 5 作品，80 年代前期作品群が 10 作品，80 年代後期作品群が 12 作品である。クラスター分析，主成分分析，判別分析，サポートベクターマシン (SVM) という 4 種類の分類法を用いた。その結果，特に教師ありの分類法である判別分析や SVM において通時的な文体変化が定量的に捉えられるようになった。とりわけ判別分析の第 1 項，第 2 項を用いた散布図において通時的な文体変化が明確に捉えられるようになった。具体的には，小谷 (1999) の「正体露見以後は作風も変化し」という主張を一部裏付けるように，正体露見以後の 1970 年代後期作品群が，正体露見以前の 1970 年代初期作品群や 1960 年代作品群とは離れた位置に布置していることが判明した。さらに，1970 年代における文体変化だけではなく，文芸批評では主張されていなかった 1960 年代作品群の独自性が判別分析の散布図から捉えることができる。また，SVM では分類正確率が 91.55%となり，サンプルサイズを考慮に入れた分類正確率の基準を有意に上回った。

本研究で用いた Alice Bradley Sheldon コーパスの作品群の中には，出版年が明らかになっているが執筆年が不明な作品が 2 作品 (*In the Great Central Library of Deneb University* と *We Who Stole The Dream*) ，出版年・執筆年ともに不明な作品が 1 作品 (*Go from me, I am one of Those Who Pall*) 含まれている。上記の研究において *In the Great Central Library of Deneb University* と *We Who Stole The Dream* に関しては出版年を基準にして分類を試み，*Go from me, I am one of Those Who Pall* に関しては分析から除外してきた。これら 3 作品に関して上記 4 種の統計手法を用いながら執筆年代の推定を探索的に試みる。

■10 月 2 日（日）
【研究発表第 6 セッション】
【研究発表 1】

英和辞典の記述とコーパスの活用

田畑　圭介（神戸親和女子大学）

紙版の英和辞典では記載できる情報量に限界があり，掲載する情報と割愛する情報を取捨選択しなければならない。Hanks(2012)は載せるべき情報と載せずにおく情報の取捨選択の根拠を大型コーパスがもたらすと述べており，取捨選択は頻度情報が基本指標となる。本発表では最初に英和辞典の記載情報が Walter(2010)の述べる識別指標，(1)written or spoken data,

(2)regional language variety，(3)synchronic versus diachronic data に基づくことを具体的な事例を見ながら考察する。次にそれぞれのデータは一つの大型コーパスから得られるわけではなく，各種のコーパスを使い分けることで，英和辞典の記述の充実につながる語彙情報が得られることを論じる。(1)については，現行のコーパスは書き言葉が多数を占めることから，独自にテレビドラマコーパスを作成し，その情報を分析することで，What do you got? ((非標準・話)) 何持っているの；(刑事ドラマなどで)何かつかんだか(話者は疑惑の念を抱いていることがしばしば暗示される)，などの口語表現の収集が可能になることを示す。(2)については autumn と fall を取り上げる。((主に英))のレーベルがつく語彙を COCA で調査することで，((主に英)) だけではない情報，*autumn* leaves [sun] 秋の紅葉[日差し](autumn は ((主に英))だが，leaves, sun, air, sky などの前では((米))でも autumn を好む傾向がある)，などが得られる。(3)については have a hard time ~ing を取り上げる。COCA，COHA，TIME コーパスを活用することで，Many people *have a hard time finding* jobs. 多くの人が就職で苦労している(いずれの時制でも time -ing とし，time to …としない;20 世紀前半までは have a hard time to do の使用がいくらか見られたが，現在ではきわめてまれ)，といった通時的な文法記述が可能となる。

英和辞典はこうした情報に加え，学習者が日常的に出会う表現の記述も求められる。北米に旅行・滞在する人であれば，インターネットで will call を利用することが予想されるが，前述のコーパスではそうした情報を抽出しにくいのが現状であり，またこうした表現を英和辞典が掲載しきれていない現状もある。日常的に出会う表現を調査する手法としての画像検索の有用性も合わせて提示する。

【研究発表 2】

The English Dialect Dictionary の原資料としての民俗学的情報の検討：特にマザーグースに注目して

谷 明信（兵庫教育大学）

本研究は *English Dialect Dictionary* (*EDD*)の原資料としての民俗学的情報の性質と範囲を検討する。その際，民俗学的情報としてマザーグース(= nursery rhymes (NR))に焦点を当て，Innsbruck 大学 Markus 教授作成の EDD online を用いて，nursery, rhyme, child(ren), Halliwell などの語や NR に特徴的な語を検索語として調査し，検討する。

Wright (1898, vol. 1: vi) は *EDD* の序で，“popular games, customs, and superstitions”の情報を丹念に収録している旨を明言する。しかし，口承により伝承されてきた文化の総体である folklore は広範であり，網羅的に調査するのは不可能であるため，本研究は NR に限定する。

辞書の原資料が問題になるのは引用が多い辞書で，研究対象はほぼ *OED2* に限られてきた。また，その研究も literary canon に集中している。一方，*EDD* の原資料の研究はほぼ行われておらず，*EDD* での NR の扱いについては不明のままである。また，辞書研究では言語的情報に焦点がおかれ，民俗学的情報等の百科辞書的情報は等閑視されがちである。

従来，*EDD* の原資料としての民俗学的情報は紙版では調査が困難であったが，Innsbruck 大学 SPEED プロジェクトによる電子化，さらに今年 4 月完成の EDD online により，以前より正確で詳細な検索が可能になり，*EDD* の原資料調査がより容易となった。

調査結果として，*OED2* と比較した場合，*EDD* の方が NR に関する情報がより多く，より詳細に記載されていることが明らかとなった。NR の詩行の引用は *EDD* の方が多く，また NR に伴うゲームの情報が記載されている場合もある。例えば，merry-ma-tanzie は *OED2* では説明されていないが，*EDD* は詳細な情報を提供する。また，網羅的な Opie（1997）でさえ記載していない NR の詩の方言でのバリエーションを記載している場合もある。原資料としての NR の考察から，*EDD* は NR の詳しい情報を記載しており，NR を重要な民俗学的情報の一つと Wright が考えていたと結論できる。

Chambers and Trudgill (2004: 102)は “[t]he study of regional variety in language can . . . be seen as one dimension of social history” と述べるが，Wright は方言とその文化が分離できないことを意

識しており，さらに *EDD* は民俗学的情報について辞書的な性質を持ち，*OED2* とは辞書としての性格が大きく異なると結論づけられる。

なお，発表では，SPEED 版と EDD online について解説し，また，その問題点についても指摘する。

■10 月 2 日（日）

【講演】

New directions in corpus linguistics: Utilizing the rich annotations found in social media data

Laurence Anthony (Waseda University)

In recent years, many researchers interested in the linguistic features of social media interactions and the impact of social media on society and culture have begun integrating corpus methods into their analyses. Twitter, Reddit, and other social media network data contain a rich set of annotations that allow researchers to identify not only the content of the message, but also the date and time it was posted, the number of 'likes' the message receives, the name of the person who posted it, their location, the names of other people in their social network, and many other details about the message. Utilizing these annotations in a corpus study allows the researcher to create a much more detailed picture and understanding of language in use than that offered by traditional corpus studies that focus mainly on Key-Word-In-Context (KWIC) and single or multi-word frequency pattern analysis.

Unfortunately, extracting data from social media networks is not a simple process. First, the researcher needs to gain access to the social media network platform through a complex authentication and verification procedure. Next, they need to navigate through the various data search and extraction protocols offered by the platform, which tend to vary from one service to another. Finally, they need to design a database architecture of their own to store the vast amount of data that can arrive from social media sites, and devise a way to view the database files that are often too large for standard text editors or corpus tools. Currently, many researchers in this area use custom scripts written in programming languages such as Python or R to carry out the complex authentication, verification, search, storage, and display processes. Although these methods are clearly successful, they create a huge barrier to researchers who do not have a strong computational background or the resources to hire a software engineer or data scientist to help them.

In this presentation, I will introduce a new corpus linguistics analysis tool called *FireAnt* that addresses many of the problems associated with social media data analysis. *FireAnt* can be used by both novice and expert computer users to access, search, store, display, and analyze social media data without the need for any programming skills. *FireAnt* includes a Twitter data collection tool and allows searching and filtering of data from various social media sites. It can also visualize that data in the form of time series plots, geolocation maps, and network graphs. In addition, the tool can output results in a variety of formats for further processing using traditional corpus tools, statistical packages, or custom scripts. In the presentation, I will briefly explain the background and motivation to develop the tool before showing how it can be used to reveal interesting linguistic features of social media interactions. I will finish the presentation with some suggestions for future ways to utilize richly annotated corpus data.

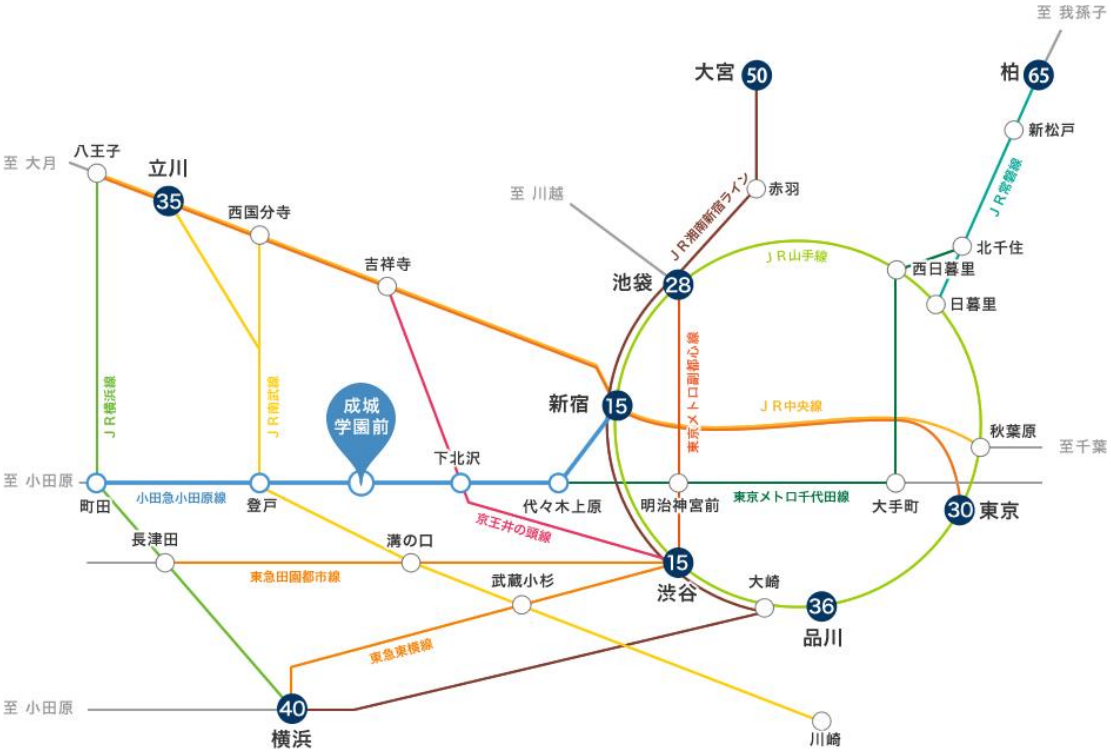
《大会参加者へのご案内》

- ワークショップの受付：会場の成城大学 7 号館 2 階 722 教室前で，午前 9 時 30 分から受付を行います。
- 大会受付：第 1 日（10 月 1 日）は成城大学 7 号館 2 階で正午から行います。第 2 日（10 月 2 日）は午前 9 時 10 分から受付いたします。
- 構内での喫煙は指定の喫煙所にてお願いいたします。
- 昼食について：第 1 日（10 月 1 日）は 7 号館地下の食堂が開店していますが，日曜日は閉店しています。近隣にコンビニ・飲食店が多数ありますので，そちらをご利用ください。
- 当日会員について：会員ではない方も，「当日会員」としてご参加いただけますので，お誘い合わせの上ご参加下さい（当日会費 2,000 円，二日間共通）。懇親会（下記）へもぜひご参加下さい。大会当日に入会受付もいたします（年会費：一般 6,000 円，学生 3,000 円）。
- 大会第 1 日の学術プログラム終了後の懇親会は，インフォーマルな雰囲気の中，参加者同士さまざまな意見交換，情報収集ができる場です。大会ご出席の方々には，ぜひ奮ってご参加いただけましたら幸いです。なお，会場準備の都合上，参加ご希望の方には事前の予約をお願いしております。ご協力のほどよろしくお願い申し上げます。
    - 英語コーパス学会第 42 回大会・懇親会
    - 日時：10 月 1 日（土）18:15–20:15
    - 場所：7 号館地下 SEIJO LOUNGE
    - 会費：5,000 円

  ※懇親会参加ご希望の方は，参加申込 Web フォーム（https://goo.gl/forms/4c095G0TDnHfOnUC3）から 9 月 22 日（木）までにお申し込み下さい。
- 懇親会の後には二次会の席も用意しております。会場確保の都合上，こちらも事前の申し込みをお願いいたします。20 名を超えた後のお申し込みにつきましては後日取り消させていただく可能性がございます。予めご了承ください。
    - 懇親会二次会
    - 日時：10 月 1 日（土）20:30–
    - 場所：JATI Seijo （http://tabelog.com/tokyo/A1318/A131814/13031766/）

  ※二次会参加ご希望の方は，参加申込 Web フォーム（https://goo.gl/forms/TEm369rsxUFpXZNw1）から 9 月 22 日（木）までにお申し込み下さい。

◆成城大学へのアクセス◆


至 我孫子
大宮 50
柏 65
新松戸
至 大月
八王子
立川 35
西国分寺
至 川越
赤羽
JR湘南新宿ライン
JR常磐線
北千住
JR山手線
吉祥寺
池袋 28
西日暮里
日暮里
JR横浜線
JR南武線
成城学園前
新宿 15
東京メトロ副都心線
JR中央線
秋葉原
至 千葉
下北沢
至 小田原
町田
小田急小田原線
登戸
代々木上原
明治神宮前
東京メトロ千代田線
大手町
30 東京
京王井の頭線
長津田
溝の口
東急田園都市線
武蔵小杉
15 渋谷
大崎
36 品川
東急東横線
至 小田原
40 横浜
川崎


成城学園前駅から成城大学まで徒歩
3分
※上記はおおよその所要時間です
成城学園初等学校
成城学園中学校高等学校
成城幼稚園
仙川
大会会場
（7 号館）
成城学園前
駅入口
正門
成城六丁目
いちょう並木
成城大学
砧八丁目
三菱東京UFJ
銀行
小田急小田原線
成城風月堂
成城石井
北口
成城学園前

◆会場案内図◆

英語コーパス学会（Japan Association for English Corpus Studies）
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　twitter: @JAECS2012　郵便振替口座:00930-3-195373　URL: http://jaecs.com/