

# 英語コーパス学会 第45回大会資料

日時：2019年10月5日（土）-10月6日（日）  
会場：高知県立大学（永国寺キャンパス）教育研究棟1階  
〒780-8515 高知市永国寺町2-22

# 英語コーパス学会 第45回大会 プログラム

## 【大会参加費】

英語コーパス学会会員：無料

非会員：当日会員参加費 2,000 円（2 日間の大会全プログラムに参加可能）

※大会・ワークショップへの事前参加予約は不要です。（ただし懇親会は事前予約制です。）

## 【展示・休憩室について】

第1日（10月5日）11:30 から、第2日（10月6日）13:30 まで、A110 教室で賛助会員等による書籍等の展示・販売を行います。展示室には茶菓もご用意いたしますので、ご休憩場所としてもご利用いただけます。

## ■第1日目

日時 2019年10月5日（土）  
受付開始 11:30（場所：教育研究棟1階）  
開会式 12:30（場所：A101）

1. 会長挨拶
2. 開催校挨拶
3. 総会
4. 学会賞審査報告
5. 事務局からの連絡

司会：石井 康毅（成城大学）  
投野 由紀夫（東京外国語大学）  
五百蔵 高浩（高知県立大学 副学長）  
西村 秀夫（三重大学）

## ●研究発表第1セッション（場所：A101） 司会：仁科 恭徳（神戸学院大学）

研究発表1：13:20–13:50

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)

研究発表2：13:55–14:25

Move Analysis of Hotel Overviews on Official Websites of Hotels in Japan: Luxury Strategies in Overview Text  
Yukie Kondo (Ritsumeikan University)

研究発表3：14:30–15:00

A Fast, Scalable, Portable Corpus Database Architecture for Small- and Large-Scale Corpus Research  
Laurence Anthony (Waseda University)

## ●研究発表第2セッション（場所：A109） 司会：菅原 崇（岐阜工業高等専門学校）

研究発表1：13:20–13:50

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）

研究発表2：13:55–14:25

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）

研究発表3：14:30–15:00

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）

<休憩 15:00–15:15>

●研究発表第3セッション（場所：A101） 司会：杉森 直樹（立命館大学）

研究発表 4：15:15–15:45

Discourse functions of high-frequency phrase frames in argumentative essays: A learner corpus study Joe Geluso (Iowa State University, PhD Candidate)

研究発表 5：15:50–16:20

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 (Kanawaza University)

●研究発表第4セッション（場所：A109） 司会：内田 諭（九州大学）

研究発表 4：15:15–15:45

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 後藤 克己（中部大学大学院生）  
—クラスター分析—

研究発表 5：15:50–16:20

Helsinki Corpus ME3-EModE3 での binomials の史的変遷 谷 明信（兵庫教育大学）  
—特に頻度と語源から—

<休憩 16:20–16:30>

●講演 16:30–18:00（場所：A101）

Revisiting A Corpus—the BNC 2014

司会：投野 由紀夫（東京外国語大学）  
Tony McEnery (Lancaster University)

<懇親会 18:20–20:20（場所：生協食堂）>

※懇親会は事前予約制です。最終ページをご覧ください。

■第2日目

日時 2019年10月6日(日)  
受付開始 9:45(場所:教育研究棟1階)

●研究発表第5セッション(場所:A101) 司会:木山 直毅(北九州市立大学)

研究発表1:10:00-10:30

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

研究発表2:10:35-11:05

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)

●研究発表第6セッション(場所:A109) 司会:能登原 祥之(同志社大学)

研究発表1:10:00-10:30

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)

研究発表2:10:35-11:05

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)

●ワークショップ:11:15-12:05(場所:A101)

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)

<休憩 12:05-13:30>

●シンポジウム:13:30-15:00(場所:A101)

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

Chair: Yukio Tono (Tokyo University of Foreign Studies)

Yukio Tono (Tokyo University of Foreign Studies)

Shin'ichiro Ishikawa (Kobe University)

Hitoshi Isahara (Toyohashi University of Technology)

Tony McEnery (Lancaster University)

閉会式 15:00-15:10(場所:A101)

■10月5日(土)

【研究発表第1セッション】

【研究発表1】

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

Tatsuya Ishii (Hiroshima University)

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 associated with moves as well as with the connection between moves in specialist genres such as basic medical science.

【研究発表2】

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.

### 【研究発表 3】

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](http://www.sketchengine.eu)), English-Copora.org ([www.english-corpora.org](http://www.english-corpora.org)) [formally [corpus.byu.edu](http://corpus.byu.edu)], and CQPWeb ([cqpweb.lancs.ac.uk](http://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.

■10月5日(土)  
【研究発表第2セッション】  
【研究発表1】

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)  
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 を含む表現の中核的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 を含むコンコードンスラインを抽出

(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)コンコードンスライン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。

【研究発表2】

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の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るとは考えにくい。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。

### 【研究発表 3】

#### 使役動詞 make の受動態補文に出現する原形不定詞の語法

村岡 宗一郎 (日本大学)

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。

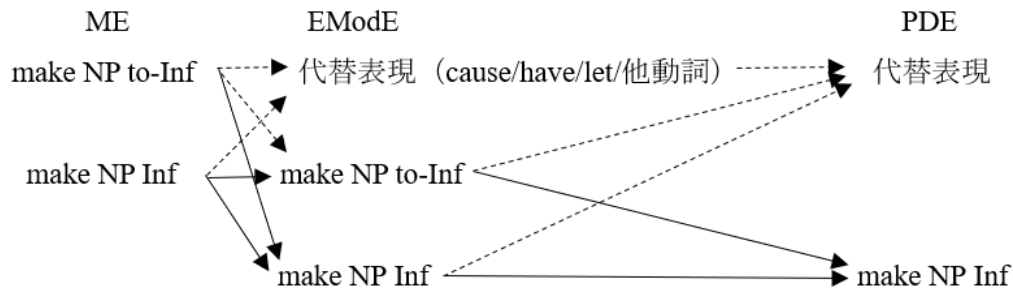
#### (2) ME における使役動詞 make と不定詞

- a. she **maketh men mysdo** many score tumes. (PPI. Biii 122)(Mustanoja (1960: 533))
- b. þe veond hit **makede 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 appear* in particular to the world, (EEBO, 1643)
- b. they *are made believe* that they are christians when they are not, (EEBO, 1658)

■10月5日(土)

【研究発表第3セッション】

【研究発表4】

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

Joe Geluso (Iowa State University)

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*, and *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.

【研究発表5】

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.

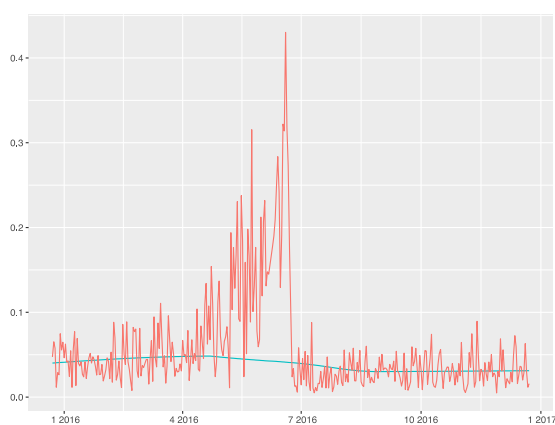


Figure 1: A topic identified via LDA for the word *Brexit* (label: “Campaign for the Brexit referendum”)

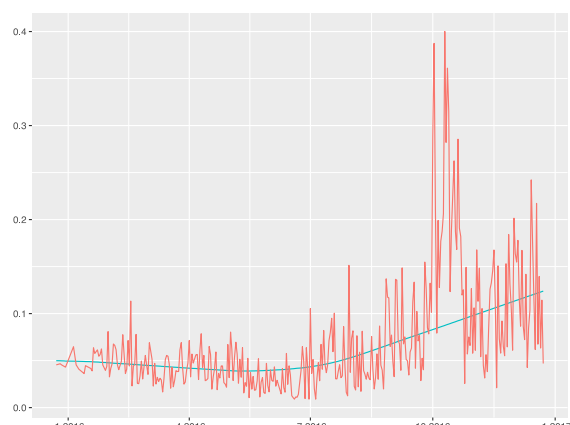


Figure 2: A topic identified via LDA for the word *referendum* (label: “Economic uncertainty associated with the referendum results”)

後藤 克己 (中部大学)

作品に生起する語彙の頻度をクラスター分析して著者推定・作品区分する場合、目的に応じて対象作品群をセクションに分割し、各語彙の生起数のセクション間におけるバラつきを集約して得られる"距離"によってクラスタリングする。そのバラつきは一般に高頻度語ほど大きく、また作品には少数の頻度上位語が際立って多く生起する特徴がある。そのためユークリッド法で距離を定義する場合、頻度上位語が偏重されることとなり、分析語数を増やしてもクラスタリングの確度は頭打ちとなる傾向がある。例えば、Dickens, Thackeray, 及び G. Eliot の各 2 作品 (計 6 作品) のコーパスを、10,000 語程度に分割 (計 62 セクション) して分析したところ、作家については各 2 作品の全セクションが上位 25 語で 1 つにクラスタリングされ正しく 3 区分されるものの、作品については分析語数を上位 1,200 語まで増やしても、6 区分中 4 区分に別作品のセクションが混在し正しく 6 区分されなかった (図 1)。区分性を向上するひとつの手段として Hoover (2003) は、分析対象を単に頻度の高い語とせず、セクション間で頻度に特徴のある語を優先する"modified method"を提案しているが、この方法でも優先した語自体の距離への影響度が高まるわけではない。

本研究は分析への低頻度語の影響度を相対的に高め、クラスタリングの確度を向上する手法を探るため、他の距離定義 (マンハッタン, キャンベラ等) を適用したデータ、並びに頻度自体を平方根, 立方根, 対数等で圧縮したデータを分析し、得られた区分性を各クラスターに含まれるセクションの不純度 (つまり著者/作品区分の不純度; Gini 係数) で評価したものである。その結果、頻度を平方根と立方根の間の圧縮となる"2.5 乗根"圧縮し、ユークリッド距離を適用した場合に最大の区分性が得られ、上記コーパスは、作家は圧縮なしの場合とほぼ同じ上位 50 語で完全に 3 区分 (Gini 係数=0) され、作品についても上位 500 語以上で殆ど完全に 6 区分された (図 2)。これは頻度上位語だけでなく、上位数百語クラスまでの語の頻度データが分析に寄与した結果であり、特に作品区分、更には同一作品内の文体変異分析の確度向上に有効な手法と考えられる。

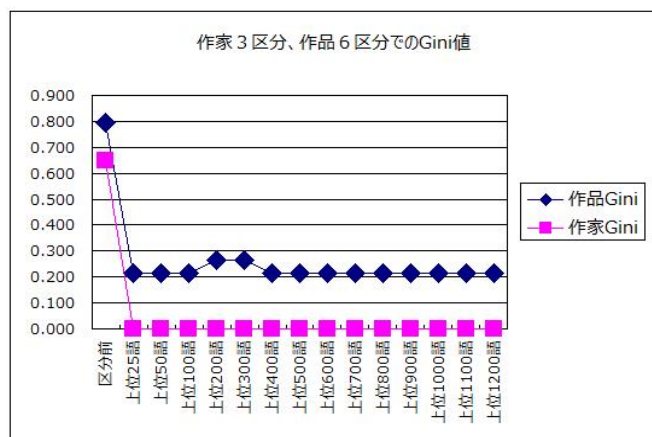


図 1 圧縮なしデータによる区分性

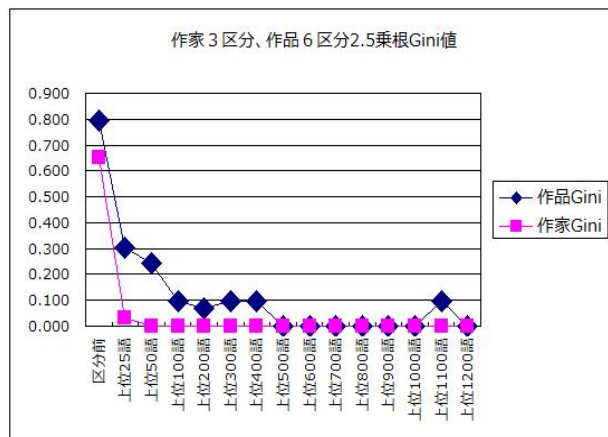


図 2 2.5 乗根圧縮データによる区分性

谷 明信 (兵庫教育大学)

近年 time and tide のような A and/or B 形式の binomials が再び注目され、後期近代英語と現代英語を調査した Mollin (2014) が、また古英語から現代英語の歴史的研究の論文を集めた Kopaczyk & Sauer (2017) が出版された。しかし、Mollin (2014; 2017)を除いて、コーパスを用いた研究は多くない。また、Helsinki Corpus を用いて調査した Pahta & Nevanlinna (1997) は appositive の一部として binomials を扱い、また、同コーパスにより調査した谷 (2017) は主に定型性の観点から調査しており、binomials 自体の史的変遷については明確な結論を出していない。

本研究は Helsinki Corpus の ME3-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 binomials の頻度と語源を調査し、その史的変遷を検討する。さらに、binomials の繰り返し度、また、構成要素間の(ir)reversibility についても調査する。また、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る。

総数 8257 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) 全体では ME4 に binomials の正規化頻度が一番高くなり、EModE1-3 にかけて漸次低下していく；
- (2) 平均では、テキストタイプ中 LAW で頻度が一番高く、次に LET NON- PRIV (公式書簡)がそれに続く；
- (3) LET NON-PRIV は ME3-4 での高頻度が EModE1 以降は半減する；
- (4) 語源の点では、ME3-4 では OE (古英語由来の語) +OE の binomials の方が OF (古仏語由来の語) +OF より割合が高いが、EModE1 からは逆転する(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)；
- (5) 語源の変遷を品詞別で見ると、名詞でのみ ME4 において OF+OF の binomials が OE+OE よりも割合が高くなり、EModE1 で同様の変化が起こる動詞と形容詞に先んじている；
- (6) 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 binomials の頻度自体が減少する EModE1 以降に、繰り返し度も減少する；
- (7) テキストタイプでは、binomials の頻度がそれほど高くない BIBLE で繰り返し度が最も高い。なお、(ir)reversibility については明確な結果は出ていない

■10月5日(土)

【講演】

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.

■10月6日(日)

【研究発表第5セッション】

【研究発表1】

「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)? get\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、より精微な調査方法が求められる。

【研究発表2】

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)

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か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。

■10月6日(日)

【研究発表第6セッション】

【研究発表1】

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 レベル別のコロケーションとそれを含むフレーズを抽出することを目的とする。具体的には、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 (表 1)。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。表 2 は(1)~(3)の具体例を示したもので、後続する括弧内は頻度を示す。

3	4	5	6
*_A key_J role_N	*_A key_J role_N *_I <sup>(1)</sup>	*_V *_A key_J role_N *_I	*_V *_A key_J role_N *_I *_A
			*_V *_A key_J role_N *_I *_V
			*_V *_V *_A key_J role_N *_I
		*_A key_J role_N *_I *_V <sup>(2)</sup>	*_V *_A key_J role_N *_I *_V <sup>(3)</sup>
		*_A key_J role_N *_I *_A	*_V *_A key_J role_N *_I *_A
	*_V *_A key_J role_N	*_V *_A key_J role_N *_I	*_V *_A key_J role_N *_I *_A
			*_V *_A key_J role_N *_I *_V
			*_V *_V *_A key_J role_N *_I
*_V *_V *_A key_J role_N		*_V *_V *_A key_J role_N *_I	

表 1 : n-gram パターン

番号	具体例
(1)	a_A key_J role_N in_I (35548), the_A key_J role_N of_I (1135), a_A key_J role_N for_I (675)...
(2)	a_A key_J role_N in_I help_V (1415), a_A key_J role_N in_I develop_V (511), a_A key_J role_N in_I ensure_V (459)...
(3)	play_V a_A key_J role_N in_I help_V (1335), play_V a_A key_J role_N in_I develop_V (475)...

表 2 : n-gram の具体例 \*A : 冠詞, I : 前置詞, J : 形容詞, N : 名詞, V : 動詞

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。



## 【研究発表 2】

特定の文法事項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(近刊)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(近刊)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を大学生のデータと比較を行った結果、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、(1)の課題ではほとんど抽出されない文法事項もあり、大学生のデータとは異なる結果も生じた。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。こ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、特定の文法項目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項目もあることが示唆され、将来的な課題である。

## 【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

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。

## 【シンポジウム】

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.

《会場アクセス》

【高知龍馬空港から】

タクシーをご利用の場合：約 40 分

バスをご利用の場合：約 50 分

空港連絡バス（とさでん交通・高知駅前観光）を利用して「北はりまや橋」（はりまや橋観光バスターミナルの次の停留所）または「JR 高知駅」までお越してください。その後のアクセスは下記の通りです。

【北はりまや橋から】

タクシーをご利用の場合：約 5 分

徒歩の場合：約 20 分

【JR 高知駅から】

タクシーをご利用の場合：約 5 分

徒歩の場合：約 20 分



《大会参加者へのご案内》

- ・会場では eduroam による無線 LAN 接続サービスが提供されています。(利用可能なのは、eduroam 参加機関のアカウントをお持ちの方に限ります。)
- ・会場には駐車場の用意はございませんので、公共の交通機関をご利用ください。
- ・大会(ワークショップを含む)への事前参加予約は不要です。ただし、懇親会(下記)への参加には予約が必要です。
- ・大会受付は、第1日(10月5日)は **11:30** から行います。第2日(10月6日)は **9:45** から行います。受付場所はいずれも教育研究棟1階です。
- ・昼食について:2日間とも学内の食堂は営業しておりませんので、近隣の飲食店等をご利用ください。繁華街としましては、徒歩10分ほどで、帯屋町(おびやまち)があります。
- ・高知県立大学文化学部の授業(「地域学実習Ⅱ」)の一環として、学生によるコンシェルジュサービスが提供されます。受付付近にて、「高知市内のまち歩き」マップ配布や、希望者への案内・説明が行われますので、観光やお食事に関する情報をお求めの際はぜひご利用ください。
- ・非会員の参加について: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(参加費2,000円;2日間の大会全プログラムに参加できます)。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。高知ならではの食事・飲み物を用意する予定ですので、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、懇親会は事前予約制となっております。参加ご希望の方は必ず事前予約をお願いします。

英語コーパス学会第45回大会懇親会

日時:10月5日(土)18:20-20:20

場所:生協食堂

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://forms.gle/F8aDBjjuHvD9p4DE9>) から 9月23日(月)までにお申し込み下さい。締め切り後の変更はできませんので予めご了承ください。

---

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](mailto:jaecs.hq@gmail.com) 郵便振替口座:00930-3-195373

URL: <http://jaecs.com/>

---