

JAECS
Japan Association for English Corpus Studies

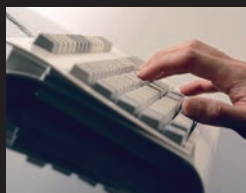
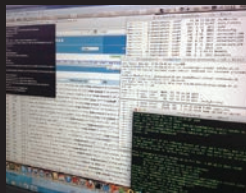
2020

英語コーパス学会

ISSN 1340-301 X

英語コーパス研究

第27号



English Corpus
Studies

27

2019年度 英語コーパス学会役員

会 長：投野由紀夫

副 会 長：石川慎一郎

事務局長：石井 康毅

理 事：新井 洋一，アントニ・ローレンス，家入 葉子，
石井 康毅，石川慎一郎，石川 保茂，岡田 毅，
金澤 俊吾，小島ますみ，小林雄一郎，杉森 直樹，
住吉 誠，滝沢 直宏，田畑 智司，中條 清美，
塚本 聡，投野由紀夫，西村 秀夫，水野 和穂

監 事：石川 有香

会 計：宇佐美裕子

顧 問：赤野 一郎，中村 純作

事務局補佐

<事務局補佐>内田 諭，大谷 直輝

編集委員会

<委 員 長>田畑 智司

<委 員>今林 修，小島ますみ，能登原祥之，水野 和穂，
水本 篤

学会賞選考委員会

<委 員 長>西村 秀夫

<委 員>五百蔵高浩，石川 保茂，高橋 薫，中條 清美，
塚本 聡

大会企画委員会

<委 員 長>金澤 俊吾

<委 員>石川 有香，小島ますみ，長 加奈子，西部真由美，
藤原 康弘，渡辺 拓人

ISSN 1340-301 X

英 語 コ ー パ ス 研 究

第 27 号

英語コーパス学会

2020

目 次

論文

日英パラレルコーパス WikipediaKyoto-LWP を用いた和英辞典の 記述改善案について—「X を固める」の場合—	仁科 恭徳	1
The Behavior of Adverbs in the Results Sections of Experimental Medical Research Articles: A Corpus-Based Move Analysis	Tatsuya ISHII and Takeshi KAWAMOTO	23
アメリカ英語における be bound to の準助動詞化の過程	家口美智子	53

研究ノート

The Expansion of the <i>how about</i> Construction in American English	Satoshi YAMAZAKI	73
---	------------------	----

資料

英語コーパス学会第45回大会資料.....		87
-----------------------	--	----

「論文」

日英パラレルコーパス WikipediaKyoto-LWP を用いた 和英辞典の記述改善案について ー「X を固める」の場合ー

仁科 恭徳

Abstract

This paper presents how a Japanese-English parallel corpus can contribute to the compilation of a Japanese-English dictionary through adopting a step-by-step, corpus-driven approach. The basic data were extracted from a general-reference corpus of Japanese, *Balanced Corpus of Contemporary Written Japanese*, as a benchmark of Japanese usage, and also from a Japanese-English parallel corpus, *The Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles*, to check the authentic Japanese-English translation units. These data are examined across semantic categories as well as descriptive grammar and frequencies, which are then compared with the descriptions present in the four current Japanese-English dictionaries. The paper highlighted an original sample manuscript based on the corpus-driven approach, to show the feasibility of utilizing parallel corpora in the bilingual dictionary compilation in the future.

1. はじめに

現在まで、日英・英日パラレルコーパスを用いた記述的・辞書学的側面からの言語研究は、仁科（2008, 2009）、染谷・赤瀬川・山岡（2011）などを除き、その数は極めて少ない。この理由は、研究者や一般ユーザーにその有用性があまり浸透しておらず、過去に一般公開された後、現在も継続して利用可能であるパラレルコーパスの数も限られていることにある。そこで、本稿では、まずパラレルコーパスの活用法の一つとして、和英辞典の記述改善のための方法論を具体的に示す。詳しくは、日本人英語学習者には英訳が難しい日本語表現「X を固める」の抽象的な意味に注目し、X 位置に生起する語の意味カテゴリーごとに、その英翻訳ユニットの量的傾向を探り、Corpus-driven Approach (Tognini-Bonelli,

2001) の見地から具体的な記述改善案を示す¹。また、その改善案と現行の 4 種の和英辞典の記述とを比較し、本稿で採用したアプローチの優位性を示す。最後に、現在構想中ではあるが数年後に一般公開を目指す一般参照日英・英日パラレルコーパスに関しても触れる。

2. 現在までの日英・英日パラレルコーパスを用いた言語研究

最近の海外のパラレルコーパス研究に目を向けると、例えば Doval & Teresa (2019) では、様々な言語間の翻訳・比較言語学研究に関するパラレルコーパスの有用性が 17 本の論文によってまとめられるなど隆盛を迎えていると言ってもよい。一方、国内の日英・英日パラレルコーパスを活用した言語記述の研究は、意味や文法、辞書学的見地から考察したものなどが見つかるが、その数は極めて限られている。代表的なものに、日英再帰形に注目した清水・村田 (2002)、身体部位を含む日英語表現を分析した田中 (2002)、when 節を取り上げた田中 (2002)、時事英語表現の翻訳傾向などを調査した仁科 (2007, 2008, 2009)、カタカナ語の誤用を取り上げた Nishina (2008)、日本語複合動詞とその翻訳を精査した染谷・赤瀬川・山岡 (2011) などがあるが、ここ 10 年近くは自然言語処理や機械翻訳などの分野以外で論文はあまり発表されておらず、依存木の統語構造的不一致から日英翻訳を分析した Oya (2017) の研究ぐらいではないだろうか。

一方、英語教育分野においては、ツール開発や DDL の教育的活用・効果検証等に関する一連の研究、例えば WebParaNews (<http://www.antlabsolutions.com/webparanews/>) の中條他 (2014, 2015)、Anthony, Chujo, & Oghigian (2011) や、SCoRE (<http://www.score-corpus.org/>) の Chujo, Oghigian, & Akasegawa (2015)、Mizumoto & Chujo (2016) などの論文が発表され研究が進んでいる。特に SCoRE に関しては、「慎重に作成した簡潔で自然な英語例文約 10,000 文と、日本人英語教師が丁寧に付けた日本語対訳文」から構成されており、英文自体のレベルが統制されている。本稿を含め日・英語の記述分析を実施する場合は、収録された英語テキストが制限や統制なしで産出されたものである必要があるため、言語学と英語教育とでは求められるパラレルコーパスの質が異なると言える。本稿ではあくまでも言語分析にもパラレルコーパスの利用価値があることを示したい。

3. 今回の分析に使用する日英パラレルコーパスとツール

今回の分析に使用した日英パラレルコーパスは、Wikipedia 日英京都関連文書対訳コーパス (<http://alaginrc.nict.go.jp/WikiCorpus/>) (以下, WK) である。染谷・赤瀬川・山岡 (2011) などによれば, 本コーパスは NICT (情報通信研究機構) が 2010 年 10 月に一般公開した Wikipedia の日本語記事 (京都関連) とその英訳から成る日英パラレルコーパスで, 2011 年 1 月の時点で開発がストップしている (最新版は ver.2.01)。京都に関する内容が中心で, 日本の伝統文化, 宗教, 歴史等の分野をカバーしている。人手翻訳による約 50 万文対を収録 (日本語の語数は約 1000 万語), 翻訳の過程 (一次翻訳→流暢さ改善のための二次翻訳→専門用語チェックの 3 段階) を記録している。

今回の分析で使用するパラレルコーパス検索ツールも, 2011 年に無料一般公開され染谷・赤瀬川・山岡 (2011) でも紹介された WikipediaKyoto-LWP (<http://wiki-kyoto.lagoinst.info/>) (以下, WKL) である。LWP は LagoWordProfiler の略で, 国立国語研究所 (以下, 国語研) と Lago 言語研究所が開発したブラウザベースのコーパス検索ツール (パルデシ・赤瀬川, 2011) を指し, 「見出し語単位で検索, コロケーションなどを文法項目に分類して整理して表示」 (染谷・赤瀬川・山岡, 2011 : p. 3) することが可能である。現在までに読売新聞とその英訳版 the Daily Yomiuri の記事から構成される日英パラレルコーパス JENAAD (the Japanese-English News Articles Alignment Database) を実装した LWP for ParaNews, 雑誌やブログなど様々なジャンルのテキストから構築された現代日本語書き言葉均衡コーパス BCCWJ (Balanced Corpus of Contemporary Written Japanese) を実装した NLB (NINJAL-LWP for BCCWJ) (パルデシ・赤瀬川, 2011) などがある。NLB では, LWP の利点を活かし日本語の動詞や名詞の共起関係や文法的振る舞いを一覧表示することが可能である。

今回の分析で WK を選定したことについて, その内容が京都関連に偏っているという課題は残るが, 現存する日英パラレルコーパスはいずれも特定ジャンルのものばかりである。むしろ, WK に収録されている訳文は上記で示した 3 つの段階を経て完成していることから, 高水準で翻訳の質が担保されている点を評価した。また, LWP を実装した WKL はパラレルコーパス上でコリゲーション検索を可能にした唯一の検索ツールである。以上の理由により, 本稿では WKL を用いることにした。

4. 先行調査

本稿では、パラレルコーパスの有効な活用法を示すために前述の WKL を用いて和英辞典編纂方法の一例を示す。調査対象には、著者が担当する翻訳の授業で過去に学生から質問があった語彙項目のうち、日本人英語学習者にとって意味の解釈や翻訳が難しいと判断した日本語動詞「固める」を選定した。

「固める」は共起する名詞と共に一つの言語ユニットとして多様な比喩的意味を表すが、この語彙項目を網羅的に調査した先行研究がほぼ皆無のため、実際にどのような意味範疇で頻繁に使われ、どのように日英翻訳される傾向にあるのかが不明である。早津 (1989) では、「固める」は有対他動詞の 1 種であり、働きかけによって生じる対象・事態・状況などの変化を含意し、動作の過程や様態よりも結果の状態に注目している動詞であるという (詳しくは、早津 (1989, 1995) を参照)。つまり「X を固める」は、X が固まった結果状態を一つのユニットと捉えることで導かれる推論の射程が広いため、X 位置に生起する名詞や文脈によって、その解釈時あるいは日英翻訳時に負荷がかかりやすい、と言えそうである。また、鍋島 (2005) は、認知言語学の視点から「防御を固める」を解釈し、POLITICS IS WAR というメタファーから「戦争の領域」→「政治の領域」へと推論が働くと指摘するが、「X を固める」にいくつかのメタファーが認められるかについては触れていない。後述のとおり各和英辞典の「固める」の記述の 70~80% 程度は比喩的意味に偏っており、掲載されている例文や表現にも辞書間でばらつきがある。このような抽象的な語彙項目やその翻訳ユニットの意味解釈および日英翻訳には、どうしても主観的な判断に頼らざるを得なかったという背景がそこにはあり、仁科 (2009) でも示したように、パラレルコーパスを活用することでそれらを「客観的・視覚的・計量的」(p. 94) に分析し記述に活かすことが可能となる。以上の理由により、日本語動詞「固める」は恰好の調査対象と判断した。

日本語辞典『大辞林 (第三版)』(2006, 三省堂)によると、日本語動詞「固める」には語義が 7 つあるが、今回はその中でも比喩的に用いられる、語義 3「不安定なものを安定させる。確実にする。」と語義 4「外部からの力に対して侵されない態勢を構える。」に該当する日英翻訳ユニットを中心に調査する。

次に事前調査として、現行の和英辞典における当該表現の記述を概観する。今回、調査対象とした辞書は、定期的且つ過去 10 年以内に一度は改訂された中型和英辞典である『ジーニアス和英辞典 (第三版)』(以下、G3)、『ウィズ

ダム和英辞典（第三版）』（以下、W3）、『オーレックス和英辞典（第二版）』（以下、O2）、『スーパー・アンカー和英辞典（第三版）』（以下、A3）の計4冊である（新和英中辞典は2002年以降、ライトハウス和英辞典は2008年以降改訂されておらず対象外、『小学館プログレッシブ』（最新は第四版）は平均改訂スピードが8.67年と他4冊よりも長いため除外）。過去には改訂に2～30年かかった和英辞典もあったが、コーパスの活用に伴い英和辞典と並んで和英辞典の改訂スピードも格段に早まっていることが表1から分かる。これら4冊の平均改訂スピードは約6.43年であり、Wシリーズの改訂スピードが約5.50年と最も短い。これは、コーパスを徹底活用するという編集方針が起因しているのかもしれない²。

表1 過去10年以内に一度は改訂された現行和英中辞典の改訂スピード

	初版	第二版	第三版	平均改訂スピード(年)
G3（語句数）	1997（80,000）	2003（82,000）	2011（83,000）	約7.00年
W3（項目数）	2007（88,000）	2012（90,000）	2018（92,000）	約5.50年
O2（用例数）	2008（80,000）	2016（98,000）		約8.00年
A3（項目数）	2000（37,000）	2004（37,000）	2012（45,000）	約6.00年

続いて、表1で挙げた現行和英辞典4種に掲載されている「固める」の記述例を表2に示す。各辞書の第一番目の語義はどれも字義通りの「<<物質>>を固める」の意であったため割愛し、全体の8割ほどを占める第二番目以降の抽象的な意の語義・用例をまとめた。

表2 現行中型和英辞典4種における「固める」(抽象的な意味)の記述比較

G3	<p>② [確かなものにする] ((決意・信念・地位などを)) confirm O; ((地位などを)) consolidate O; ((土台・基盤などを)) strengthen O 彼のアドバイスで私は決意を固めた [= 彼のアドバイスが私の決心を固めた] His advice confirmed my decision. / 業界における地位を固める consolidate one's position in the industry / チームの結束を固める strengthen the team's unity / さらなる成長の基礎を固める strengthen the foundations for further growth / もうそろそろ身を固めてもいい頃だろう It's about time you settled down.</p> <p>③ [守りを強める] 敵の攻撃に備えて町の防備を固める fortify the city against the enemy attack / 国境の守りを固める strengthen the defense of the border</p> <p>④ [構成する] 先発メンバーをベテランで固める use players full of experience as starting members</p>
W3	<p>② [[強固にする]] strengthen; [[意見・信念などを]] confirmed; [[地位などを]] ((書)) consolidate. ・組織を固める strengthen the organization. ・信念を固める strengthen one's belief. ・地位を固める consolidate one's position. ・防備 [守備] を固める strengthen the defense. ▶ 国境の守りを固めねばならない We must strengthen our defenses on [^x of] the border. ▶ 彼の手紙を読んで留学の決意を固めた His letter confirmed (me in) my decision to study abroad. / I was confirmed in my decision to study abroad by his letter.</p> <p>③ [[身を固める]] settle down. ▶ 彼はどのようにして結婚して身を固めないのか Why doesn't he marry and settle down?</p>
O2	<p>② <<しっかりさせる>> よい英文を書くにはまず文法の基礎を固める必要がある In order to write good English, you need to build a solid grammatical base. / 彼は父の仕事を継ぐ決心を固めた He [firmly resolved [OR decided firmly] to take over his father's business. / こぶしを固める clench [OR double] ONE'S fist (s) / 事業計画を固める strengthen a business plan / そろそろ君も身を固めなきゃ It's about time you settle down. (→結婚する) It's about time you got married. / 次の試合に勝つためにはもっと守備を固めないといけない In order to win the next game, you have to have a stronger defense.</p> <p>③ <<守りを厳重にする>> 国境を固める secure a national border / その建物は武装した警官によって固められた That building was secured by armed police.</p>
A3	<p>② 【安定させる】 consolidate+㊦(強固にする); strengthen+㊦(強化する) ▶ 首相は実力者を閣僚に起用して政権を固めた The prime minister consolidated his power by picking influential politicians as a cabinet ministers. 私は会社を辞める決意を固めた I made up my mind to quit the company. 中井さんは20代に事業の基礎を固めた Nakai built up the foundations of his business in his twenties. きみもそろそろ身を固める年だ It is time you started thinking about marriage. / You are old enough to get married and settle down. ▶ ジャイアンツは9回の裏守備を固めた The Giants strengthened their defense in the bottom of the ninth.</p> <p>③ 【まとめる】 ▶ みんなの荷物は固めて1か所に置いておきなさい Gather your things together and put them in one place. その政治家の証言はうそで固めたものだった The politician's testimony was nothing but a pack of lies.</p>

基本的にはほとんどの用例がヲ格を用いた「Xを固める」の用法を掲載していた。他の辞書の記述と比較して G3 と W3 の良い点は、翻訳ユニットごとに X (の意味) が (()) や [[]] で顕在化されていることにある。ただし、

G3 の記述で不親切なのは、X が「地位」の場合に *confirm* と *consolidate* の二通りが示されていること、((決意・信念・地位)) が同じカテゴリーで括られていること(「地位」に違和感を感じる)、語義3「守り」は語義2に含めてもよく(())でXを顕在化させるべきであること、語義4[構成する]はその頻度から掲載しなくてもよいことが指摘できる。W3に関しては、*confirm* や *consolidate* 同様に *strengthen* と共起しやすい意味や名詞も顕在化させた方がユーザーフレンドリーである。また、「信念」の場合は *confirm* を使うことを明示しているにもかかわらず、用例は *strengthen one's belief* となっている。Xの意味カテゴリーごとに用例を差し込む方が分かりやすいのではないか。この点も考慮して後に独自の記述改善案を示す。

これら4種の和英辞典の用例として掲載された「X を固める」の翻訳ユニットをX位置の名詞ごとに原形でまとめたものが表3である。

表3 各和英辞典が掲載している「X を固める」の翻訳ユニット

G3 (7種)	決意・決心 <i>confirm one's decision</i> , 地位 <i>consolidate one's position</i> , (チームの) 結束 <i>strengthen the (team's) unity</i> , 基礎 <i>strengthen the foundations</i> , 身 <i>settle down</i> , 町の防備 <i>fortify the city</i> , 守り <i>strengthen the defense</i>
W3 (7種)	組織 <i>strengthen the organization</i> , 信念 <i>strengthen one's belief</i> , 地位 <i>consolidate one's position</i> , 防備・守備 <i>strengthen the defense</i> , 守り <i>strengthen one's defenses</i> , 決意 <i>confirm one's decision/be confirmed in one's position</i> , 身 <i>settle down</i>
O2 (7種)	基礎 <i>build a solid ~ base</i> , 決心 <i>firmly resolve/decide firmly</i> , こぶし <i>clench or double one's fist(s)</i> , 事業計画 <i>strengthen a business plan</i> , 身 <i>settle down or get married</i> , (もっと) 守備 <i>have a strong(er) defense</i> , 国境 <i>secure a national border</i>
A3 (5種)	政権 <i>consolidate one's power</i> , 決意 <i>make up one's mind</i> , 基礎 <i>build up the foundations</i> , 身 <i>think about marriage/get married and settle down</i> , 守備 <i>strengthen one's defense</i>

G3では、「決意・決心」など固める対象が<心>の翻訳ユニットを最優先に掲載し、<防御>の意の用例は最後に配置されている。一方、W3では「信念」と「決意」が離れて配置されている。O2の「決心」の翻訳ユニットはG3のものとは異なり、O2の「守備」の翻訳ユニットはW3と異なる。A3の「決意」と「基礎」の翻訳ユニットはG3のものと違って口語により近いものである。「固める」の翻訳等価である英語動詞に注目すると、G3は *confirm*, *consolidate*, *strengthen*, *fortify* の4種、W3は *strengthen*, *consolidate*, *confirm* の3種、O2は

build, resolve(decide), clench(double), strengthen, have, secure の 6 種, A3 は *consolidate, make up, build up, strengthen* の 4 種を掲載している。どの辞書も翻訳ユニットの内容と掲載順序に統一性が見られず、意味や表現がコーパス・データに基づいて頻度順になっているのか疑問が残る。

5. 結果と考察

現在のところ、一般参照日英・英日パラレルコーパスは存在していない。そこで、本節では、まず日本語の母集団をある程度反映しているであろう NLB で日本語の調査を行い、和英辞典に掲載すべき項目・用例を精査した上で日英パラレルコーパスの WKL から翻訳ユニットを抽出し、その特徴と傾向をまとめる。そして、そのデータに基づき Corpus-driven Approach の見地から和英辞典の記述改善案を具体的に示す。

5.1 X を固める

WKL を用いた調査に先立ち、NLB (ver.1.40) で「固める」を検索した結果、NLB に生じた全 24,719 種の動詞の中で「固める」は頻度順で第 938 位 (1551 例) であった。また、「固める」の全用例中の約 70% (1084 例) がヲ格の名詞 + 動詞コリゲーション「X を固める」で用いられていた。よって、表 2 の和英辞典の「固める」の項でヲ格の用例が主に掲載されていたことは妥当であろう。なお、X 位置に生じた名詞は計 320 種であった。

続いて、WKL で「固める」を検索した結果、生じた全 10,968 種の動詞の中で「固める」は頻度順で第 418 位 (608 例) であった。また、名詞 + 動詞コリゲーション全 363 例中、ヲ格の「X を固める」は 54.27% (197 例) を占め最多であった。なお、X 位置に生じた名詞は計 64 種であった。

表 4 は、これら 2 種のコーパスにおいて、X 位置に生じた名詞を頻度・MI スコア・LD スコア順に上位 10 種をまとめたものである³。一般参照コーパス (NLB) と特定ジャンルコーパス (WKL) というレジスターの違いから、各統計指標において X 位置に生起している名詞の傾向にある程度の乖離がみられることが分かる。

表4 「X を固める」目的語 Top10

	NLB（320 種）						WKL（64 種）					
順位	頻度		MI	LD		頻度	MI	LD				
1	決意	103	地歩	15.15	決意	10.37	守り	17	戦備	13.75	守り	11.01
2	身	102	ベークライト	13.81	守り	9.60	基盤	15	守り	13.64	防備	10.52
3	方針	49	本管	13.16	決心	9.00	地位	15	防備	13.18	決意	10.14
4	守り	35	守り	13.13	地歩	8.94	基礎	13	身持ち	13.16	基盤	9.54
5	基礎	30	傍証	13.13	結束	8.50	防備	12	糯米	12.94	基礎	9.27
6	決心	27	塔屋	13.03	足場	8.49	体制	11	決意	12.53	防御	8.92
7	脇	21	臍	13.03	防備	8.28	決意	10	足場	12.05	方針	8.78
8	基盤	19	決意	12.79	方針	7.98	方針	10	結束	11.35	意志	8.40
9	体制	18	四つ辻	12.71	拳	7.86	身	6	控え	11.19	地位	8.34
10	意思	18	結束	12.45	守備	7.72	防御	5	土台	11.16	後ろ	8.24

Note. MI = Mutual Information score, LD = Log Dice score

次に、NLB から抽出された全 320 種と WKL からの全 64 種をリスト化し、各名詞に意味カテゴリーを付与した。意味カテゴリー付与には、国立国語研究所コーパス開発センターが公開中の分類語彙表-増強改訂版データベース (https://pj.ninjal.ac.jp/corpus_center/goihyo.html) を参照した。結果、該当する部門数は NLB が 6 種（活動、関係、自然、主体、生産品、生産物）と WKL が 5 種（活動、関係、自然、主体、生産物）、中項目数は NLB が 38 種と WKL が 23 種であった⁴。図 1 は X 位置に生起した名詞の部門別の割合を示す。

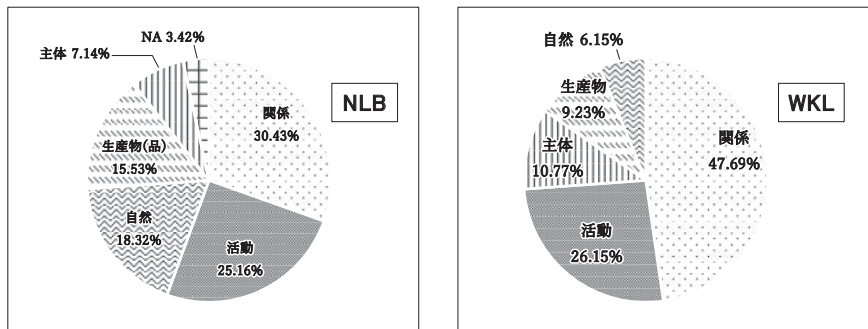


図1 分類語彙表（部門）に基づく X の意味カテゴリー（割合）

図1のNLBの検索結果から、一般的に関係や活動の意味を示す名詞がX位置に生じやすいことが分かる。また、WKLの検索結果との比較から、割合順位の第三位以降はコーパス間で意味カテゴリーの種類が異なることも分かる。「固める」の字義通りの意における対象は「自然」や「生産物」であり、比喩的な意における対象は「関係」や「活動」である。全体的な割合は後者の方が圧倒的に大きいため、頻度主義の立場から言えば、和英辞典には抽象的意味→字義的意味の順で訳語や用例を掲載すべきなのかもしれない。

図2は、NLBにおいてX位置に生じた名詞を中項目ごとに分類し頻度順に並べ替えたものである。この結果から、他の項目よりも多用されている総頻度50回を超過した心、身体、空間、交わり、類の5種の意味カテゴリーに属する名詞を調査対象とする。なお、空間と類は関係部門に、心と交わりは活動部門に、身体は自然部門に属している。

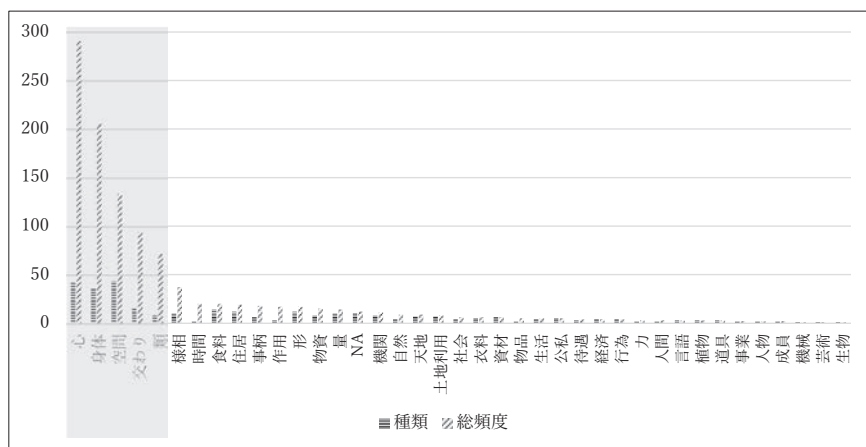


図2 NLBの分類語彙表(中項目)に基づくXの意味カテゴリー(種類数と総頻度)

表5は、図2の結果から選定された5種の意味カテゴリーに属する名詞に関して、WKLで再検索した結果を示す。次節以降、表5に基づきWKLを用いて当該表現の翻訳傾向を調査するが、意味カテゴリー「活動-心」をMIND、「自然-身体」をBODY、「関係-空間」をSPACE、「活動-交わり」をDEFENSE、「関係-類」をPOSITIONと呼ぶことにする。特に、「SPACEを固める」と「DEFENSEを固める」に生じた名詞の特性から、互いに幾許かの意味的な関連性が認められそうであり、それらの翻訳ユニットに関しても何かしらのオーバーラップ

が生じる可能性は否めない。しかしながら、前者はあくまで部分的で具体性を帯びた名詞(例. 国境)が生起する一方で、後者は全体的且つ抽象的な名詞(例. 守り)が生起していることから、本稿ではこの点に人為的な手を加えず、分類語彙表に従って別の意味カテゴリーとして捉え、データ駆動的に分析を進めることとする⁵。

表5 WKL から抽出された調査対象5種の意味カテゴリーとその名詞

意味カテゴリー (種類 / 総頻度)	WKL の X 位置に生起した名詞 (頻度)
活動－心 (7 種 / 計 28 回)	方針 (10), 決意 (10), 意志 (3), 心 (2), 志 (1), 意思 (1), 覚悟 (1)
自然－身体 (2 種 / 計 10 回)	身 (6), 脇 (4)
関係－空間 (16 種 / 計 22 回)	周り (3), 後ろ (3), 周囲 (2), 足場 (2), 北方 (1), 南側 (1), 国内 (1), 国内外 (1), 国境 (1), 地盤 (1), 屋外 (1), 拠点 (1), 東側 (1), 根元 (1), 穂先 (1), 要地 (1)
活動－交わり (5 種 / 計 38 回)	守り (17), 防備 (12), 防御 (5), 守備 (3), 戦備 (1)
関係－類 (3 種 / 計 43 回)	地位 (15), 基盤 (15), 基礎 (13)

5.2 「MIND を固める」の翻訳ユニット

WKL において、意味カテゴリー MIND に属する名詞は計 7 種である。「MIND を固める」の翻訳ユニットを WKL で調べた結果、総頻度順に *decide (firmly) to do* (方針 4, 決意 1, 意志 1, 心 2, 覚悟 1), *make up one's mind to do* (方針 1, 決意 2), *decide on (a policy of ~ing)* (方針 1, 決意 1), *determine to do* (決意 2), *establish a policy of ~ing* (方針 1), *determine the policy for ~ing* (方針 1), *settle on the ~ policy* (方針 1), *it v-link decided that S + V ~* (方針 1), *make one's decision* (決意 1), *strengthen determination of N* (決意 1), *make a resolve* (決意 1), *make decision to do* (決意 1), *strengthen one's intention to do* (意志 1), *N's intention to do* (意志 1), *make sure of one's ambition for ~ing* (志 1), *the stiffening of the resolve* (意思 1) であった。結果、*decide (firmly) to do* が 5 種の日本語名詞に共通して使われる翻訳ユニットであり全体の 32.14% を占めていた。また、固い決意や意思などを示す場合は副詞 *firmly* が伴うことも分かった。尚、「方針」

の場合は名詞 *policy*, 「意志」の場合は名詞 *intention* が使われる傾向にあることも分かった。

5.3 「BODY を固める」の翻訳ユニット

WKLにおいて、意味カテゴリー BODY に属する名詞は計2種である。「身を固める」の翻訳ユニットは *wear* (2例), *get married* (1例), *be in* (1例), *with* (1例), *put* (1例) の5種であった。G3やW3のように *settle down* のみ掲載している辞書も多いが、O2やA3にも掲載されている「結婚する」の意の *get married*, や「<衣服で>身を固める」の意の *wear* の用例を和英辞典には優先的に掲載すべではないだろうか。また、「脇を固める」は全用例において *support* が用いられていた(例: *he played important role supporting leading players*- 人名 PNM03188)。

5.4 「SPACE を固める」の翻訳ユニット

WKLにおいて、意味カテゴリー SPACE に属する名詞は計16種である。ただし、他の意味カテゴリーと比較して、どの名詞も低頻度である。動詞に注目した翻訳ユニットの内訳(頻度順)は、*reinforce* + PLACE (後ろ1, 南側1, 国内外1, 東側1) (例: *Hiroshige, who was reinforcing the east side of the city* - 歴史 HST00033), *guard* + place (周り1, 後ろ1, 周囲1) (例: *He then belonged to Kondo's squad and guarded the perimeter of Ikedaya* - 人名 PNM02539), *strengthen* N (周り1, 地盤1), *consolidate* N (足場1, 要地1), *secure* + place (国境1, 拠点1), *surrounded by* N (周り1), *make place safer* (後ろ1), *place people around* N (周囲1), *solidify one's own base* (足場1), *defend* + place (北方1), LT/paraphrase (国内1, 根元1), *shore up* + place (屋外1), *firm the N tip* (穂先1)であった。総頻度やレンジの結果から、*reinforce* や *guard* を掲載し、分かり易い例文を作例するのが妥当ではないか。

5.5 「DEFENSE を固める」の翻訳ユニット

WKLにおいて、意味カテゴリー DEFENSE に属する名詞は計5種である。DEFENSE は総頻度が他の意味カテゴリーより多いため、翻訳ユニットを動詞ごとにまとめて示す。表6から、計7種の動詞 (*strengthen*, *reinforce*, *build up*, *fortify*, *protect*, *guard*, *solidify*) を用いて訳出されており、中でも *strengthen* と *reinforce* は特筆して使用されていることが分かる。others (その他) の計4例は、

いずれも具体的な動詞で置き換えるというよりも, *increase one's defenses*, *defense v-link so strong*, *for protection* など, やや意識的に訳出されていた。

表6 「DEFENSEを固める」の翻訳ユニット

		<i>strengthen</i>	<i>reinforce</i>	<i>build up</i>	<i>fortify</i>	<i>others</i>	<i>protect</i>	<i>guard</i>	<i>solidify</i>	<i>LT</i>
守り	頻度	8	2	2	0	2	1	1	0	1
(17)	%	47.06	11.76	11.76	0.00	11.76	5.88	5.88	0.00	5.88
防備	頻度	2	3	2	4	0	0	0	1	0
(12)	%	16.67	25.00	16.67	33.33	0.00	0.00	0.00	8.33	0.00
防御	頻度	2	2	0	0	1	0	0	0	0
(5)	%	40.00	40.00	0.00	0.00	20.00	0.00	0.00	0.00	0.00
守備	頻度	0	3	0	0	0	0	0	0	0
(3)	%	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
戦備	頻度	0	0	0	0	1	0	0	0	0
(1)	%	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
計	頻度	12	10	4	4	4	1	1	1	1
(38)	%	31.58	26.32	10.53	10.53	10.53	2.63	2.63	2.63	2.63

Note. LT = Lost in Translation

まず, 「守りを固める」全17例のうち (FQ: 17, MI: 13.64, LD: 11.01), 1語で翻訳されていたのは *guard* と *protect* のわずか2例 (11.76%) で, LTの1例を除いた他14例は2語以上で翻訳されており, その6例 (35.29%) は目的語が複数形であった。動詞 *strengthen* の8例中4例は, *strengthen one's defense(s)* で翻訳されていた。また, *others* (その他) の例として, *increase one's defenses*, *defense v-link so strong* の翻訳表現が見つかった。

次に, 「防備を固める」全12例のうち (FQ: 12, MI: 13.18, LD: 10.52), 1語で翻訳されていたのは *fortify* 全4例中の1例のみで (例 . *It is also described that the Dazaifu fortified the area with the hills behind after the Toi invasion in 1019 - 歴史HST00745*), 他11例は2語以上で訳出されていた。また, 全体の6例 (50.00%) は目的語が複数形で, *reinforce* の1例を除いて, 他の全用例は目的語位置に *defense(s)* が生起していた。

そして、「防御を固める」は5例 (FQ: 5, MI: 10.92, LD: 8.92) のみで, *strengthen* と *reinforce* は共に *defense(s)* を目的語にとっていた。その他の例として, *for protection* で訳出されている例が1例見つかった。これら以外にも, W3 や O2 に掲載されている「守備を固める」は3例のみ (FQ: 3, MI: 10.15, LD: 8.17) 見つかり, そのいずれも *reinforce* が使われていた (*reinforce the defense(s) / defensive ability*)。「戦備を固める」は逐語訳ではなく *prepared for war* があてられていた。

以上から, X が「守り」の場合は *strengthen one's defense(s)*, 「防備」の場合は *fortify the defense(s)*, 「防御」と「守備」の場合は *reinforce the defense(s)* と棲み分けることができる。

5.6 「POSITION を固める」の翻訳ユニット

WKL において, 意味カテゴリー POSITION に属する名詞は計3種である。このうち, 「地位」と他2種は意味的に異なるため別個に分析する。まず, 「地位を固める」全15例の翻訳ユニットを調査した結果, その内訳は頻度順に *solidify one's position/status (as N)* (6例), *strengthen one's position as N* (2例), *build up a presence/position (as N)* (2例), *reinforce one's position as N* (1例), *cement one's position as N* (1例), *become established as N* (1例), *one's position as N is consolidated* (1例), *succeed by Ving* (1例) であった。動詞は全体の40%を占める *solidify* が優勢であり, 翻訳ユニットのほとんどが「V one's position as 職業・役割」のパターンで使われていた。

次に, 表7では他2種の名詞「基盤」と「基礎」の翻訳ユニットを動詞ごとにまとめた。具体的な英語動詞として *strengthen*, *establish*, *solidify*, *build*, *form*, *develop*, *become* の計7種が確認された。LT に関しては1件見つかった(例. *[T]he base of the clan, eventually caused a struggle among the Amago clan in destroying the Ouchi clan, and also the Rebellion of Okihisa pushed the Amago family into a quagmire.*- 人名 PNM02605)。

表7 「基盤・基礎を固める」の翻訳ユニット

		<i>strengthen</i>	<i>establish</i>	<i>solidify</i>	<i>build</i>	<i>form</i>	<i>develop</i>	<i>become</i>	<i>LT</i>
基盤	頻度	8	3	2	0	0	0	1	1
(15)	%	53.33	20.00	13.33	0.00	0.00	0.00	6.67	6.67
基礎	頻度	4	2	2	3	1	1	0	0
(13)	%	30.77	15.38	15.38	23.08	7.69	7.69	0.00	0.00
計	頻度	12	5	4	3	1	1	1	1
(28)	%	42.86	17.86	14.29	10.71	3.57	3.57	3.57	3.57

Note. LT = Lost in Translation

全体として *strengthen* が 12 例（全体の 42.86%）と最も多く用いられており、そのうち 9 例（75.00%）が目的語に *foundation* をとっていた。他にも *establish*, *solidify*, *build* が少なからず用いられることも分かった。また、全 28 例を調べた結果、LT の 1 例を除いた 27 例が「動詞＋目的語」の共起表現で訳出されており（1 例のみ受動文）、目的語には *foundation(s)*（18 例、66.67%）あるいは *base/basis*（9 例、32.14%）のいずれかで、ほぼ全用例で前置詞句が後続し、その 75.00%（21 例）が *of* 句であった。また、「固める」を直訳していない例として、LT（1 例）以外に *become* と *develop* の例が見つかった。前者は「基盤を固める」の翻訳ユニットとして *become the foundation of N*（旧家 FML00070）、後者は「基礎を固める」の翻訳ユニットとして *develop a solid base of N*（歴史 HST00825）であった。

5.7 和英辞典における「固める」の記述改善案

前節までの NLB と WKL の分析結果から、和英辞典における抽象的な意の「固める」の記述改善案を以下に示す。

★ [方針・決意などを] *decide, make up one's mind* ▶ ケンは留学する**決意を固めた** Ken *decided firmly [OR made up his mind] to study abroad.* ★ [(衣服で) 身を] *wear* ▶ マイクは黒っぽいスーツで**身を固めて**オペラを観に行った Mike *went to see an opera wearing a dark suit.* / [(結婚して) 身を] *get married* ▶ メアリーは**身を固める**決心がついていない Mary *is not ready to get married.* / [脇を] *support* ▶ ミアは主役の**脇を固める**重要な役割を担った Mia *played an important role supporting a leading player.* ★ [周り・背後などを] *reinforce, guard* ▶ 警察はアメリカ大使館の**周りを固めて**いた The police *guarded [OR reinforced] the area around the U.S. embassy.* ★ [守り・防備などを] *strengthen, fortify* ▶ 我々は自国の**防備を固める**必要がある We need to *strengthen [OR fortify] the defense of our own nation.* ★ [地位などを] *solidify* ▶ ジムは芸術家としての**地位を固めた** Jim *solidified his position as an artist.* / [基盤・基礎などを] *establish, build* ▶ ニュートンは古典物理学の**基盤を固めた** Newton *established the basis of classical physics.* ▶ スティーブ・ジョブスはアップル社の**基礎を固めた**男である Steve Jobs *is the man who built the foundations of Apple Inc.*

まず、意味カテゴリー MIND からは、多頻度名詞「方針」や「決意」がバランスよく翻訳された *decide (firmly) to do* や *make up one's mind to do* を採用すべきであろう。次に、意味カテゴリー BODY からは、「身」の字義的意味は他動詞 *wear* を、非字義的意味は *get married* を、「脇」は他動詞 *support* を掲載すべきであろう。また、意味カテゴリー SPACE からは、*reinforce/guard + PLACE* を採用すべきであろう。そして、意味カテゴリー DEFENSE からは、多頻度である「守り」と「防備」を取り上げ、前者の翻訳ユニットは *strengthen the/one's defense(s)*、後者は *fortify the defense(s) of N* を採用すればどうか⁶。頻度(割合)・レンジを考慮すれば他動詞 *reinforce* も候補に挙がるが、意味カテゴリー SPACE の箇所既出のため、別の訳語を採用した方がユーザーフレンドリーである。最後に、意味カテゴリー POSITION は「地位」と「基盤・基礎」に分け、前者は *solidify one's position/status as N* を、後者は *strengthen the foundation(s) of N* を採用すればどうか。ただし、意味カテゴリー DEFENSE の欄で *strengthen* は既出のため、「基盤」には *establish*、「基礎」には *build* を掲載するのも方策である。尚、*establish* は目的語に *basis* を取りやすい。

上記の記述改善案はあくまで試案ではあるが、翻訳の専門家が扱う翻訳ユニットを掲載することで、現行の和英辞典よりもオーセンティックな仕上がりになっている。例えば、W3 の「決意を固める」の例文(彼の手紙を読んで留

学の決意を固めた *His letter confirmed (me in) my decision to study abroad. / I was confirmed in my decision to study abroad by his letter.*) よりも、パラレルコーパスのデータに基づいた上記の改善案の方が自然であろう。

6. 結語：一般参照日英・英日パラレルコーパスの構築に向けて

本稿では、現在使用可能である日英パラレルコーパスを用いることで可能となる和英辞典編纂論の一端を示した。現時点では一般参照日英・英日パラレルコーパスが存在していないため、本稿ではその代替案として NLB や WKL を徹底的に活用した。WK に収録されているテキストは京都関連の新聞であることからジャンルが偏っている。また「翻訳者の数も限られていることから、特定の翻訳者の言語的嗜好や癖が平均化されないまま残っている可能性」(染谷・赤瀬川・山岡, 2011 : p. 14) もある。よって、抽出したデータと現行辞書の記述を直接比較することは避けた方がよいのかもしれないが、それでも、特定の日本語表現に対応するオーセンティックな英翻訳ユニットの抽出には十二分に活用できる。

現状、一言語コーパスのように、収録テキストの種類やサイズ、比率などを考慮したパラレルコーパスは存在していないため、今後の課題として、辞書編纂にも援用できる大規模一般参照日英・英日パラレルコーパスを構築し、比較言語学的分析や Linguae (<https://www.linguae.jp/>) のような翻訳支援ツールの詳細・発展版として教育的・実用的に使える検索システムを開発したい。特に、学生が和英翻訳時に参照するツールには、アルク社の英辞郎 on the WEB (<https://www.alc.co.jp/>) (英辞郎 on the WEB Pro は 2019 年 5 月現在で和英見出し 356 万収録) や Weblilo 英語例文検索 (<https://ejje.weblilo.jp/>, 以下 Weblilo) (2019 年 6 月 8 日時点で 986 万例文収録) など web ベースのものが多く。前者は、元々は特定の個人による翻訳をまとめたもので誤訳も混ざっている。後者は、プレミアム会員になることで情報源の選択も可能となるが問題点もある。例えば、Weblilo で「言い放つ」を検索すると計 8 例がヒットするが、6 例は WK から 2 例は EDR 日英対訳辞書からの用例である。つまり一般ユーザーが知りたい翻訳表現を検索した場合、WK という特殊なジャンルの翻訳データのみを鵜呑みにする可能性が高い (Weblilo に搭載されている翻訳データの中で WK の比重が大きい可能性もある)。また、WKL で当該表現を再検索してみると、実際には 9 例の用例がヒットしたため、Weblilo 検索では WK の全用例が反映されて

いない可能性もある。

そこで、英語学習者のためにもまずは既存の日英・英日パラレルコーパスのフォーマットを整備し、既存のパラレルコーパスコンコーダンサーはもちろん、LWPのようなワード・プロファイラーに全パラレルコーパスを搭載し一覧検索できるシステム開発が望まれる。例えば、既存の日英・英日パラレルコーパスは、アライメントの方法が文あるいは段落単位とコーパスによって異なる。また、生テキストのものやXMLでタグ付けされたものもある。さらに、言語ごとに別ファイルでまとめられているものもあれば、同一ファイルにまとめられているものもある。そこで、全てを統一のフォーマットで再整備する必要がある。現状、構想中の擬似一般参照日英・英日パラレルコーパスでは、複雑な検索にも対応できるように、アノテーション（品詞付与、構文解析）を施し、XMLフォーマットで整える予定である。

次に、翻訳方向、ジャンル、サイズ等のリソース情報を整備・参照することで、他のジャンルとのバランスや比率を考慮し、不足しているジャンルの日英・英日パラレルコーパスを新たに構築する必要がある。例えば、仁科（2008）の交換可能性に関する調査結果から、英日と日英は別物として翻訳方向ごとに整備・構築する方がよい。また、BNCのように特定の時代の言語テキストのみを収集してパラレルコーパスを構築することは翻訳物という特性から資料収集が難しいため、せめて2000年以降の文書や発話など大枠のルールを設定した上でジャンル構成を考えれば、本コーパスの構築も夢ではない。特に、話し言葉の日英・英日パラレルコーパスの種類が現状では限られているため、映画やドラマ、TEDの字幕データをパラレルコーパス化するなど対策を考えたい⁷。

参考までに、パラレルコーパスの素材になりえるネット上の記事が、日英パラレルニュース記事URL集（http://www.jaist.ac.jp/project/NLP_Portal/doc/par_news/readme.html）に紹介されており、CNET Japan（7759記事対）、WIRED.jp（14098記事対）、EIC ネット（1739記事対）、NHK ニュースで英会話（950記事対）、NHKABC ニュースシャワー（444記事対）に加え、Wall Street Journal（有料）などがある。ただし、これらの翻訳には「抄訳が多い、翻訳の質」といった問題点もある（私信：赤瀬川史郎氏 2019年6月6日）。元データがウェブの場合はコーパス化するのは容易であるが、紙媒体の場合はOCRを用いる必要がある。山岡・赤瀬川パラレルコーパスプロジェクトでは複数のOCRソフトを用いて読み取りデータの誤りを補正したが、英語と比較して日本語の読み取り精度は低いため、テキストファイル化が容易な Kindle

本など電子書籍も活用したい。

これらの下地があれば、SL と TL の両言語からのコリゲーション検索、検索した語や句、表現のジャンル別の割合、抽出した翻訳ユニットのコリゲーションの一覧表示（頻度・割合情報含む）などの機能も検索システムに搭載することが可能となり、これからの比較言語学研究的の有用なツールとなりうる。また、信頼性の高い翻訳支援ツールとしても活用可能であろう。ただし、このようなパラレルコーパスを搭載した検索システム構築には時間・コスト・技術のいずれもが必要となる。Sketch Engine (<https://www.sketchengine.eu/>) もパラレルコーパスの分析が可能であることから、こちらを使うのも一つの方策かもしれない（私信：赤瀬川史郎氏 2019 年 6 月 6 日）。

謝 辞

本稿の執筆に際し、大変貴重なご助言とご示唆を賜りました査読委員 3 名の先生方に心より感謝申し上げます。また、WikipediaKyoto-LWP の利用をご快諾くださった Lago 言語研究所の赤瀬川史郎所長にもここに感謝の意を記します。

注

1. Tognini-Bonelli (2001) は Corpus-driven Approach を提唱し、既に存在していた collocation (コロケーション) と colligation (コリゲーション) の概念に加えて、semantic preference (優先的意味選択) と semantic prosody (意味的韻律) という新しい言語事実を取り上げた。McEnery, Xiao, & Tono (2006) にも手際よい解説があるが、collocation は語と語が共起する関係、colligation は語と文法カテゴリーとの関係（あるいは文法カテゴリー間の関係、詳しくは Firth, 1968 を参照）、semantic preference は語と意味的に類似した語群との関係、semantic prosody は態度や評価といった感情的な意味を生み出す共起関係のことを指し、順に抽象性が増していく。
2. 各和英辞典では、改訂の度に斬新な特色がプラスされている。例えば、O2 では、アイデアの泉、英作文のヒント、PLANET BOARD、How to Say It など独自コラムを設けて特色を明確に出しており、G3 の Communication Box では「目的→英語」と言語機能ごとに英語表現をまとめている。最近では、大学受験やビジネスを意識した構成にもなっている。
3. LD スコアはコーパスサイズの影響を受けないため、異なるコーパス間でスコアの比較が可能となる (Rychlý, 2008; 参照 https://www.sketchengine.eu/my_keywords/logdice/)。
4. 分類語彙表に掲載されていなかった名詞はまとめて NA で示している。ただし、例えばウォーターは水と判断できるように、掲載されていなくても意味として同等と判断できるものに関しては筆者の判断で類別し各部門・中項目にカウントして

いる。

5. この点に関して、査読者の方から貴重なご指摘を頂いた。ここに感謝の意を示す。
6. 2019年8月17日付で「動詞（現・過）＋名詞（単・複）」のGoogleドメイン検索を実施。例えば strengthen(ed) ～ defense(s) の場合、現×単が421,000件（英）／842,000件（米）、現×複は78,400件（英）／117,000件（米）、過×単が122,000件（英）／366,000件（米）、過×複が15,500件（英）／61,700件（米）であったので、現×単の例文を採択。他の用例に関しても同様の調査を実施。
7. 映画やTEDには台本・原稿が存在しているため、厳密には自然言語の話し言葉ではないという意見もある。また、字幕の著作権の問題もあることから、構築後に無償公開できるかは現時点では未定である。

参考文献

- Anthony, L., Chujo K. and Oghigian, K. (2011) “A novel, web-based, parallel concordancer for use in the ESL/EFL classroom”. In Newman, J., Baayen, H. and Rice, S. (eds.), *Corpus-based Studies in Language Use, Language Learning, and Language Documentation*. Amsterdam: Rodopi, pp. 123–138.
- 中條清美・アントニローレンス・内山将夫・西垣知佳子 (2014) 「フリーウェア WebParaNews オンライン・コンコーダンスの英語授業における活用」『日本大学生産工学部研究報告B』第47号：49–63.
- 中條清美・西垣知佳子・赤瀬川史郎・内山将夫 (2015) 「レキシカル・プロファイリング型オンラインコーパス検索ツール LWP for ParaNews の英語授業における利用」『日本大学生産工学部研究報告B』第48号：45–57.
- Chujo, K., Oghigian, K. and Akasegawa, S. (2015) “A corpus and grammatical browsing system for remedial EFL learners”. In Leńko-Szymańska, A. and Boulton, A. (eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, pp. 109–128.
- Doval, I. and Teresa, S. N. (2019) *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications (Studies in Corpus Linguistics)*. Amsterdam: John Benjamins.
- Firth, J. R. (1968) *Selected Papers of J. R. Firth 1952–59*. London: Longman.
- 早津恵美子 (1989) 「有対他動詞と無対他動詞の違いについて－意味的な特徴を中心に－」『言語研究』第95号，231–256.
- 早津恵美子 (1995) 「有対他動詞と無対他動詞について－意味的な特徴を中心に－」早津恵美子・須賀一好（編）『動詞の自他』ひつじ書房，pp. 179–197.
- 岸野英治 (2018) 『ウィズダム和英辞典（第三版）』東京：三省堂．
- McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- 南出康世・中邑光男 編 (2011) 『ジーニアス和英辞典（第三版）』東京：大修館書店.
- Mizumoto, A. and Chujo, K. (2016) “Who is Data-Driven Learning for? Challenging the Monolithic View of its Relationship with Learning Styles.” *System* 61, 55–64.

- 鍋島弘治郎 (2005) 「批判的ディスコース分析と認知言語学の接点—認知メタファー理論の CDA への応用—」『時事英語学研究』第 44 号：43-55.
- 仁科恭徳 (2007) 「相互関係を表す形容詞から見たシノニム学習の理論と実践教材：実証的考察とパラレルコーパスを用いたデータ駆動型学習法を中心に」『LET 関西支部研究集録』第 11 号：45-59.
- 仁科恭徳 (2008) 「パラレルコーパスを用いた交換可能性の一考察」『英語コーパス研究』第 15 号：81-95.
- Nishina, Y. (2008) “Parallel corpora in computer-assisted language learning: A case of lexical studies and data-driven learning using moodle”. In Marriott, R. and Torres, P. (eds.), *Handbook of Research on E-Learning Methodologies for Language Acquisition*. Hershey: Information Science, pp. 203-217.
- 仁科恭徳 (2009) 「パラレルコーパスを用いた抽象語彙・フレーズの一考察：これからの二言語辞書の編纂論」『LET 関西支部研究集録』第 12 号：83-97.
- 野村恵造ほか 編 (2016) 『オーレックス和英辞典 (第二版)』東京：旺文社.
- Oya, M. (2017) “Syntactic Divergence Patterns among English Translations of Japanese One-Word Sentences in a Parallel Corpus.” *English Corpus Studies* 24, 19-40.
- プラシャントパルデシ・赤瀬川史朗 (2011) 「BCCWJ を活用した基本動詞ハンドブック作成コーパスブラウジングシステム NINJAL-LWP の特長と機能」, 特定領域研究「日本語コーパス」総括班. 『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集.
- Rychlý, P. (2008) “A Lexicographer-Friendly Association Score.” *Proceedings of Recent Advances in Slavonic Natural Language Processing*, 6-9.
- 清水眞・村田真樹 (2002) 「パラレルコーパスを用いた日英再帰形の分析」『英語コーパス研究』第 9 号：17-34.
- 染谷泰正・赤瀬川史朗・山岡洋一 (2011) 「大規模翻訳コーパスの構築とその研究および教育上の可能性」『日本メディア英語学会第 1 回年次大会発表資料』1-15.
- 田中啓 (2002) 「「顔」を含む日本語表現と対応する英語表現について」『英語コーパス研究』第 9 号：57-80.
- 田中美和子 (2002) 「『語り』の when 節」の意味特徴」『英語コーパス研究』第 9 号：81-92.
- Tognini-Bonelli, T. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- 山岸勝榮 (2015) 『スーパー・アンカー和英辞典 (第三版)』東京：学研.

(仁科 恭徳 神戸学院大学 E-mail: ynishina@gc.kobegakuin.ac.jp)

「論文」

The Behavior of Adverbs in the Results Sections of Experimental Medical Research Articles: A Corpus-Based Move Analysis

Tatsuya ISHII and Takeshi KAWAMOTO

Abstract

Move analysis has clarified the functions of moves and steps, which the writers in a discourse community share, and corpus linguistics has developed the concept of phraseology as a new unit of analysis. These two types of studies can be integrated in investigations using corpora based on move analysis. Creating corpora based on move analysis can lead to the description of lexical phrases strongly associated with the functions of steps. Previous studies on this topic have been conducted mainly in the clinical medicine field within the framework of the Introduction, Methods, Results, and Discussion (IMRD or IMRAD) structure by calculating n-grams or using keyword analysis. However, the behavior of 4-grams with adverbs has not yet been investigated. Furthermore, keyword analysis has almost exclusively focused on the behavior of a limited range of grammatical words, neglecting the analysis of other types of words such as adverbs. Thus, in this study, we created a corpus based on the move analysis of experimental medical research articles, which included a total of approximately 1.6 million words. The Results sections of the studied articles included three moves: (RM1) Introducing experiments, (RM2) Announcing results, and (RM3) Commenting results. Identifying 4-grams with adverbs revealed the overall word combinations, although some low-frequency adverbs were not extracted in 4-grams. To describe the behavior of low-frequency adverbs and place the word combinations into lexical phrases, we observed the concordance lines of adverbs in each move while referring to the Life Science Dictionary. In conclusion, we found 26 lexical phrases strongly associated with the functions of three steps in RM1, three steps in RM2, and two steps in RM3. The current study opens the possibility of using a corpus based on move analysis to establish lexical phrases strongly relevant to the functions of steps for use in teaching English for specific purposes.

1. Introduction

1.1 Move Analysis and Phraseological Studies

Swales (1990: 140) proposed the “Create a Research Space (CARS) model,” advocating the functions of “moves” and “steps” and providing a brief list of expressions, especially for the Introduction sections. Dudley-Evans and St John (1998: 89) defined a “move” as “a unit that relates both to the writer’s purpose and to the content that s/he wishes to communicate” and a “step” as “a lower level text unit than move that provides a detailed perspective on the options open to the writers.” According to Coffin (2001) and Hyland (2002), move analysis has uncovered the functions of moves and steps shared by the writers in a “discourse community”—a group of people that “has a broadly agreed set of common public goals” (Swales, 1990: 24). Move analysis of research articles (RAs) in a particular discourse community reveals the functions of moves and steps in the traditional Introduction, Methods, Results, and Discussion (IMRD or IMRAD) structure, wherein these are included as article sections. As Thompson (1993) and Williams (1999) pointed out, move analysis mainly focuses on the Introduction and Discussion sections rather than the Results sections. However, Swales (1990: 175–176) stated that “the major differences do not lie so much in Introductions and Discussions (where I believe most people expect it) but rather in the Method and Results sections.” Thus, this study focuses on the functions of moves and steps in the Results sections of RAs.

Regarding the move analysis of medical RAs, Nwogu (1997) investigated 15 clinical medical RAs and identified 11 moves, including two moves in the Results sections. Clinical medical RAs “apply an initial hypothesis to a largely heterogeneous population; data collection is followed by definition, analysis and conclusions,” while “experimental research is based on hypothesis, a largely controlled method and observation, followed by deduction and further hypothesizing, perhaps leading to more experimentation” (Williams, 1996: 185). In fact, Williams (1996) found different and common uses of verbs in two clinical and experimental medical RAs. In addition, Thompson (1993) and Williams (1999) reported the rhetorical patterns of experimental medical RAs and biomedical RAs, respectively. Similarly, Kanoksilapatham (2005, 2007) identified the flow of discourse in experimental biochemical journals. Although these previous studies on move analysis illustrated moves or steps different from those

of clinical medical RAs, they did not exhibit the list of expressions involved. Thus, the current study attempts to reveal the linguistic features associated with the functions of moves and steps in the Results sections of experimental medical RAs.

In contrast with move analysis, corpus linguistics provides novel insights for analyzing language from the viewpoints of lexis and grammar as continua while developing the concept of phraseology as a new unit of analysis (Sinclair, 2004). Previous phraseological studies calculated n-grams or conducted keyword analysis to reveal the “recurrent patterns of associated words” (Cheng, 2012: 101) in a variety of labels and definitions, such as “lexical phrases,” “formulas,” “routines,” “fixed expressions,” “prefabricated patterns” (or “prefabs”), “n-grams,” and “lexical bundles” (Gray & Biber, 2015: 125). In this study, “n-grams” are defined as word combinations calculated by special software, CasualConc (Imao, 2019), while “lexical phrases” are “the general construct of phraseological patterns of three or more words,” (Gray & Biber, 2015: 127) generalized by investigating the concordance lines.

As a pioneering study combining move analysis and phraseological studies, Mizumoto et al. (2016) built a corpus based on move analysis. They collected and conducted move analysis on around 1,000 RAs in the area of applied linguistics published during 2000–2015 for a total of approximately 85 million words. They created a web-based system for supporting academic writing by exhibiting typical 3-gram to 5-gram units for each move and their concordance lines in the area of applied linguistics as well as of computer science, materials science, and medicine.¹ Regarding the field of clinical medical RAs, Saber (2012) created a clinical medial corpus based on IMRD as four subcorpora, collecting 375 articles published during 2006–2009 that included approximately 1.2 million words. Because “keywords can be lexical items which reflect the topic of a particular text” (Hunston, 2002: 68), he established keyword lists for each subcorpus by comparison with the whole corpus. Moreover, he reported 3-grams to 5-grams with nouns or verbs in the keywords and identified the functions of steps. However, little is known about the functions of n-grams with adverbs in RAs. Thus, as a first step of this study, using a corpus based on move analysis, we examined 4-grams with adverbs in keywords in the Results sections of experimental medical RAs. As Hyland (2008a: 8) advocated, 4-grams are “far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles.”

Although n-grams are produced automatically by using special software such as Antconc (Anthony, 2019), the “frequency-driven approach” (Biber, Conrad, & Cortes, 2004: 1) has been criticized from the perspective of the method of identification. When words such as adverbs and adjectives are inserted, the word sequences are regarded as indicating different phraseological patterns (Durrant, 2009); Namely, some phraseological patterns might be ignored (Cheng, 2012). To reduce the possibility of this issue, keyword analysis has mainly been carried out by observing the concordance lines carefully (Hunston, 2002). For instance, in the area of medical RAs, Kawamoto and Ishii (2018a, 2018b) investigated a corpus of approximately 1.4 million words consisting of 12 moves, modified by Nwogu (1997), from 395 clinical medical RAs published during 2013–2014. The keywords in each move were determined according to the value of the log-likelihood ratio (LLR) with comparison to the whole corpus. Moreover, Kawamoto and Ishii (2018a), following Gledhill (2000a, 2000b), observed the concordance lines of grammatical words such as *has*, *that*, and *our* in the Introduction and Discussion sections. However, little is known about how adverbs in keywords behave in RAs. Moreover, reading concordance lines of adverbs leads to the identification of lexical phrases with adverbs that are not evident due to their low frequency. Here, as a second step of this study, we identified lexical phrases with adverbs associated with the functions of steps in the Results sections of experimental medical RAs.

1.2 Functions of Moves and Steps in the Results Sections of Experimental Medical RAs

To build a corpus based on move analysis of experimental medical RAs, the moves need to be identified for comparison with those in previous studies. Although Thompson (1993) analyzed 36 Results sections of experimental medical RAs and found six rhetorical moves, the steps were not noted at all. On the other hand, based on Brett (1994), Williams (1999: 347) investigated eight Results sections of biomedical RAs and identified “three broad classes,” which can be regarded as moves, although the term *categories* was used. Moreover, the functions of *categories*, which can be identified as steps, were reported as in Table 1.

Table 1: Functions of the Classes (Moves) and Categories (Steps) in Results Sections Identified by Williams (1999: 363–364)

Classes	Categories
Metatextual Categories	0.1 Pointer 0.2 Structure of Section
Presentational Categories	0.3 Procedural 1.1 Statement of Findings 2.1 Substantiation of Finding 2.2 Non-validation of Finding
Comment Categories	3.0 Explanation of Finding 3.1 Comparison of Finding with Literature 3.2 Evaluation of Finding or Hypothesis 3.4 Implications of Finding

In addition, Kanoksilapatham (2005: 290–291) collected 60 biochemical RAs comprising 0.32 million words to identify 15 distinct moves, including four moves in the Results sections: “Stating procedures,” “Justifying procedures or methodology,” “Stating results,” and “Stating comments on the results.” Although the same corpus data were analyzed, Kanoksilapatham (2007) renamed the four moves in Results sections as Move 8 to Move 11 and described the functions of their steps, as presented in Table 2.

Table 2: Functions of the Moves and Steps in Results Sections Identified by Kanoksilapatham (2007: 76)

Move 8:	Step 1: Describing aims and purposes
Restating methodological issues	Step 2: Stating research questions Step 3: Making hypotheses Step 4: Listing procedures or methodological techniques
Move 9: Justifying methodological issues	
Move 10:	Step 1: Reporting results
Announcing results	Step 2: Substantiating results Step 3: Invalidating results
Move 11:	Step 1: Explaining results
Commenting results	Step 2: Generalizing/interpreting results Step 3: Evaluating results Step 4: Stating limitations Step 5: Summarizing

Tables 1 and 2 include some common functions with different names or definitions. For instance, Kanoksilapatham (2007) did not identify “Metatextual Categories” as shown in Table 1, while “0.3 Procedural” in Table 1 seems equivalent to “Move 8: Restating methodological issues” and “Move 9: Justifying methodological issues” in Table 2. Thus, in this study, “Move 8: Restating methodological issues” and “Move 9: Justifying methodological issues” in Table 2 were combined as the first move called (RM1) Introducing experiments. Following Kanoksilapatham (2007), the second move was labeled as (RM2) Announcing results, and the third move was designated as (RM3) Commenting results. As for RM2, “1.1 Statement of Findings” in Table 1 can be regarded as “Step 1: Reporting results” as part of “Move 10: Announcing results” in Table 2, while “2.1 Substantiation of Finding” and “2.2 Non-validation of Finding” in Table 1 can be considered as “Step 2: Substantiating results” and “Step 3: Invalidating results” as part of “Move 10: Announcing results.” Regarding RM3, “Comment Categories” in Table 1 functioned as “Move 11: Commenting results” in Table 2.

In sum, given the functions of moves and steps in Tables 1 and 2, in this study, the texts were segmented into three moves: (RM1) Introducing experiments, (RM2) Announcing results, and (RM3) Commenting results (see Table 3), and tagged RM1, RM2, and, RM3, respectively.

2. Methods

2.1 Overview

To describe the behavior of adverbs, we followed the “top-down corpus-based analysis” advocated by Biber, Connor, and Upton (2007: 13). To apply a top-down approach, we collected 304 articles from 30 leading journals (identified in the Appendix) utilizing the IMRD structure and created a corpus according to the move analysis.² When conducting move analysis manually, we consulted the text independently with reference to the criteria presented in Section 1.2 and identified the boundaries between the moves (see Section 2.2). First, to follow a corpus-based approach, using Antconc (Anthony, 2019), we identified keywords according to the value of LLR with comparison to the whole corpus (approximately 1.6 million words) as a reference corpus. TagAnt (Anthony, 2015) was used to check parts of speech in context and count the frequencies of words that were used as adverbs in RM1, RM2,

and RM3. When the keywords behaved as more than one part of speech, we focused on the behavior of the keywords as adverbs. For instance, the words *next*, *first*, and *further* in RM1 and *overall* in RM2 and RM3 worked as not only adverbs but also adjectives. Similarly, when the adverbs functioned in an adverbial phrase, we investigated the behavior of keywords as adverbs and excluded the analysis of the adverbial phrases. For example, in RM3, although the adverbial phrases *taken together* and *all together* occurred 174 times and six times, respectively, we described the behavior of the adverb *together*. Next, to reveal frequent word combinations, based on CasualConc (Imao, 2019), 4-grams with adverbs in keywords in each move were extracted according to the frequency criteria of 20 times per million words (Hyland, 2008b). The minimal frequencies for 4-grams in RM1 and RM2 were determined to be five and eight, respectively. Although the frequency in RM3 should be two according to the calculation, the corpus size of RM3 was so small that it was considered to be at least four. In calculating 4-grams, the adverb *not* was not considered in RM2 and RM3. Although the meanings of phrases with and without the adverb *not* are opposite, both such phrases were identified. For example, the phrase *was significantly reduced in* and the phrase *was not significantly reduced in* can be considered equivalent.

To observe the behavior of some adverbs that were not extracted in 4-grams and to describe lexical phrases with adverbs, reading the concordance lines of adverbs in keywords was conducted, with reference to their raw frequency and LLR. To check the word combinations and place them into lexical phrases, we consulted the Life Science Dictionary (LSD) available on the web.³ The LSD is a monitor corpus that consists of about 100 million words from abstracts published in about 150 leading journals during 1998–2015 and freely available from PubMed.⁴ Because “it seems to be the case that most abstracts reflect the IMRD pattern of the RA itself” (Swales, 1990: 181), using the LSD corpus was an effective method for checking the word combinations and the behavior of adverbs. Lastly, we categorized the lexical phrases with adverbs according to their functions and simultaneously labeled the names of the steps based on Kanoksilapatham (2007), as shown in Table 2.

2.2 Criteria for the Boundaries Between Moves

Although Kanoksilapatham (2005, 2007) reported move structure and example sentences, she did not present clear criteria for boundaries between moves. To maintain

replicability, our corpus data were created according to the concept that tense and reporting verbs function as signals of move segments (Waard & Henk, 2012). Tense denotes authors' rhetorical purposes: The past tense refers to specific experiments, while the present tense generalizes the findings (Malcolm, 1987). Thus, the past tense indicates the signals of RM1 and RM2, whereas the present tense functions as the signal of RM3 (see Table 3). The choice of the reporting verbs relates to the rhetorical function. The reporting verbs included "Procedural Verbs" (*examined*, *used*, and *evaluated*), which served as the signal of RM1, "Objective Verbs" (*found* and *observed*), which functioned as the signal of RM2, and "Post-Experimental Verbs" (*suggest* and *indicate*), which indicated RM3 (Thomas & Thomas, 1994: 134–139). Using verb-based analysis, we segmented all texts manually. For example, when two signal verbs appeared in one sentence, we separated the sentence into two different moves. Accordingly, we identified the signals of RM1 and RM2 in the following sentence.

- (1) We **examined** expression of [.....] **and found** that astrocytes at [.....].
(Tasdemir-Yilmaz & Freeman, 2014: 22 [emphases added])

Furthermore, a participial construction such as *suggesting* led us to divide a sentence into RM2 and RM3, as in the following sentence.

- (2) Notably, these ratios **did not change** between [.....], **suggesting** that the [.....].
(Case et al., 2014: 868 [emphases added])

2.3 Corpus Data of the Results Sections

In creating a corpus, one difficulty is in how to treat balance and representativeness (Biber, 1990, 1993). In this study, corpus data were obtained from the compiled articles with the IMRD structure based on the three criteria advocated by Nwogu (1997: 121): "representativity, reputation, and accessibility." First, to guarantee the representativeness of the corpus data, the span of journals was limited, as Saber (2012) mentioned. Second, regarding the reputation, leading journals in the field of experimental medicine, including the four journals analyzed by Kanoksilapatham (2005, 2007), were selected using the Journal Impact Factor (Journal Citation Reports) and the number of citations. Lastly, regarding accessibility, all articles were extracted

from e-journals (available from the Hiroshima University Library) that could be transferred easily from HTML pages to Notepad on a personal computer. The figure legends, tables, headings, and section titles were excluded. The corpus consisted of 304 articles from 30 leading experimental medical journals published in 2014 (approximately 1.6 million words). Table 3 presents the overall information and the functions of moves in the Results sections.

Table 3: Corpus Data on the Results Sections

	Tokens	Number of files	Moves	Applied Signals
RM1	243,711	4,398	Introducing experiments	Procedural Verbs past tense (examined / used / evaluated)
RM2	404,428	5,379	Announcing results	Objective Verbs past tense (found / observed)
RM3	84,281	3,137	Commenting results	Post-Experimental Verbs present tense / participial construction (suggest / indicate suggesting / indicating)
Total	732,420	12,914		

3. Results

3.1 Overview of Adverbs and 4-grams with Adverbs in the Results Sections

Table 4 shows the adverbs identified as keywords and their raw frequencies in the three moves in the Results sections. For RM1, RM2, and RM3, the numbers of keywords identified were 94, 249, and 108, while the numbers of adverbs in the keywords were 5, 27, and 12, respectively.

Table 4: Adverbs as Keywords of the Results Sections

RM1	next (533), further (289), therefore (274), first (244), stably (48)
RM2	not (2,198), also (1,353), significantly (1,026), only (721), however (649), interestingly (252), furthermore (233), moreover (222), notably (219), similarly (216), indeed (206), alone (195), importantly (178), strongly (141), approximately (134), completely (118), almost (115), slightly (98), markedly (93), consistently (91), nearly (81), substantially (80), dramatically (76), remarkably (67), strikingly (66), conversely (66), surprisingly (62)
RM3	not (483), thus (404), together (216), therefore (153), collectively (97), rather (66), directly (66), strongly (57), hence (38), overall (33), functionally (23), altogether (16)

To investigate the frequent word combinations shown in Table 4, 4-grams with adverbs were produced. Table 5 shows 4-grams and their frequencies with adverbs. These 4-grams occurred at least five times. Hereafter, the adverbs listed in Table 4 are written in *italics*.

Table 5: 4-Grams with Adverbs and Their Frequencies in RM1

<i>we next</i> sought to (40)	<i>we next</i> examined whether (8)
<i>we next</i> examined the (27)	<i>we next</i> assessed whether (8)
<i>we next</i> asked whether (21)	<i>next we</i> examined the (8)
<i>we next</i> investigated whether (20)	<i>next we</i> asked whether (8)
<i>we next</i> investigated the (17)	<i>we therefore</i> hypothesized that (7)
<i>to further</i> investigate the (17)	<i>we therefore</i> examined the (7)
<i>we next</i> tested whether (16)	<i>we next</i> analyzed the (7)
<i>we next</i> tested the (15)	<i>to further</i> examine the (7)
<i>we next</i> determined whether (15)	<i>to further</i> test the (7)
<i>we first</i> examined the (12)	<i>we therefore</i> tested whether (6)
<i>next</i> sought to identify (11)	<i>further</i> investigate the role (6)
<i>we next</i> assessed the (9)	<i>to further</i> examine whether (5)
<i>we first</i> tested the (9)	<i>next we</i> tested whether (5)
<i>to further</i> explore the (9)	<i>therefore</i> we examined the (5)
<i>to further</i> confirm the (9)	<i>we next</i> compared the (5)
<i>to further</i> characterize the (9)	<i>we next</i> explored the (5)
<i>next</i> sought to determine (9)	<i>we next</i> used the (5)

These 4-grams with adverbs in Table 5 are strongly associated with the function of (RM1) Introducing experiments. The frequent word combinations were the pronoun *we* with the adverbs and the adverb *further* between *to*-infinitive and some verb stems. However, other word combinations could be made but may not have been extracted.

Table 6 shows the 4-grams with adverbs in RM2 that were found at least eight times.

Table 6: 4-Grams with Adverbs and Their Frequencies in RM2

we <i>also</i> found that (16)	we <i>also</i> observed that (10)
was <i>also</i> observed in (15)	found to be <i>significantly</i> (9)
was <i>significantly</i> higher in (15)	<i>notably</i> we found that (9)
were <i>significantly</i> enriched in (14)	<i>interestingly</i> we found that (8)
<i>however</i> we found that (13)	<i>however</i> in contrast to (8)
was <i>significantly</i> reduced in (12)	were <i>also</i> observed in (8)
we <i>also</i> observed a (11)	<i>significantly</i> more likely to (8)
was <i>significantly</i> lower in (11)	was <i>significantly</i> higher than (8)
were <i>significantly</i> reduced in (11)	<i>alone</i> or in combination (8)
<i>also</i> observed in the (10)	

In Table 6, the 4-grams with the adverbs shown above were used to realize the function of (RM2) Announcing results. For example, the adverbs *notably* and *interestingly* appeared with the pronoun *we* and were used to emphasize important results. On the other hand, the adverb *significantly* was used with the passive voice to indicate an increase or decrease. Although 7 of 27 adverbs related to RM2, which were listed in Table 4, were identified by calculating 4-grams, there was a need to investigate the behavior of the other 20 adverbs by observing the concordance lines carefully. Table 6 shows 4-grams with adverbs in RM3. The listed 4-grams appeared at least four times.

Table 7: 4-Grams with Adverbs and Their Frequencies in RM3

<i>together</i> these results indicate (15)	<i>together</i> these findings indicate (5)
<i>together</i> these data suggest (14)	<i>together</i> these results suggest (5)
<i>together</i> these data indicate (12)	<i>collectively</i> these results indicate (5)
<i>collectively</i> these data suggest (10)	<i>altogether</i> these results demonstrate (4)
<i>collectively</i> these results suggest (8)	<i>collectively</i> these data indicate (4)
<i>thus</i> we conclude that (7)	<i>collectively</i> these results demonstrate (4)
<i>we therefore</i> conclude that (7)	results <i>strongly</i> suggest that (4)
<i>together</i> these data demonstrate (7)	<i>thus</i> indicating that these (4)
<i>together</i> these data suggested (6)	<i>together</i> these data indicated (4)
<i>together</i> these results demonstrate (6)	<i>we thus</i> conclude that (4)

In Table 7, these 4-grams with adverbs were strongly relevant to the function of (RM3) Commenting results. The frequent word combinations were the nouns *results*, *data*, and *findings* with the adverbs *together*, *collectively*, and *altogether*. Moreover, the pronoun *we* appeared with the adverbs *thus* and *therefore*, and the verb *conclude*. However, some adverbs in RM3 of Table 4 had not yet been investigated due to their low frequency.

3.2 Behavior of Adverbs in RM1

3.2.1 Relationship Between Adverbs and Steps in RM1

As mentioned in Section 2.1, to observe the behavior of some adverbs that were not be extracted in 4-grams due to their low frequency and to describe lexical phrases with adverbs, the concordance lines of adverbs in keywords were carefully observed. Moreover, the LSD was consulted to check the word combinations and place them into lexical phrases. The lexical phrases were categorized according to their functions, and the names of steps were labeled. In (RM1) Introducing experiments, three steps were identified: Step (1) “Describing aims and purposes,” Step (2) “Making hypotheses,” and Step (3) “Listing procedures or methodological techniques,” based on Kanoksilapatham (2007: 76). As presented in Table 8, five adverbs in keywords were extracted using the value of LLR, nominalized frequency (NF) calculated per 1,000 words, and raw frequency. The rank was determined according to the value of LLR.

Table 8: Adverbs in RM1

Rank	Keyword	LLR	NF	Frequency	Step
1	next	695.87	2.19	533	Steps (1) and (3)
2	first	84.90	1.00	244	Steps (1) and (3)
3	therefore	80.24	1.12	274	Steps (1), (2), and (3)
4	further	70.08	1.19	289	Steps (1) and Step (3)
5	stably	27.29	0.20	48	Step (3)

Note: LLR = log-likelihood ratio; NF = normalized frequency.

As shown in Section 3.1, the adverbs *next*, *first*, *therefore*, and *further* tended to co-occur with the pronoun *we*. For instance, for the adverb *next*, the phrase *we next* occurred 391 times while the phrase *Next, we* appeared 124 times.

3.2.2 RM1-Step (1): Describing Aims and Purposes

To indicate Step (1) Describing aims and purposes, the pronoun *we* appeared with verbs related to what was planned in the study, such as *asked* and *determined*, followed by, as an option, either the conjunction *whether* or *if* or the noun phrase *the role* or *the effect(s)* (Table 9). Hereafter, vocabulary items in the same column are interchangeable, while those separated by a horizontal line are not. The word combinations were checked by referring to the LSD.

Table 9: Lexical Phrase (1) in RM1-Step (1)

We	<i>next</i> <i>first</i> <i>therefore</i> <i>further</i>	asked	whether if
		addressed analyzed assessed determined evaluated examined explored investigated tested	whether if the role of the effect(s) of
<i>Next,</i> <i>First,</i> <i>Therefore,</i>	we	characterized	the role of the effect(s) of
		compared	the effect(s) of

As an alternative choice for describing aims and purposes, the pronoun *we* emerged with the verbs followed by *to*-infinitive (Table 10).

Table 10: Lexical Phrase (2) in RM1-Step (1)

We	<i>next</i> <i>first</i> <i>therefore</i>	attempted sought set out wanted	to
<i>Next,</i> <i>First,</i> <i>Therefore,</i>	we		

As shown in Section 3.1, the adverb *further* appeared with *to*-infinitive in the phrase *To further* 176 times. The phrase *To further* with several typical verbs was related to consolidating the given findings (Table 11).

Table 11: Lexical Phrase (3) in RM1-Step (1)

<i>To further</i>	confirm establish validate verify
-------------------	--

On the other hand, the phrase *To further* with several typical verbs was also used to indicate a further exploration, as shown in Table 12.

Table 12: Lexical Phrase (4) in RM1-Step (1)

<i>To further</i>	address assess characterize define delineate demonstrate determine dissect elucidate evaluate examine explore investigate narrow down probe quantify test understand
-------------------	---

3.2.3 RM1-Step (2): Making Hypotheses

As the second step in RM1, the adverb *therefore* leads to an account of developing the hypotheses. The verb *hypothesized* with the adverb *therefore* was observed by producing 4-grams with the keyword. Furthermore, observing the concordance lines of the adverb *therefore* revealed similar word combinations with the verb *reasoned*, as shown in Table 13.

Table 13: Lexical Phrase in RM1-Step (2)

we	<i>therefore</i>	hypothesized reasoned	that
----	------------------	--------------------------	------

3.2.4 RM1-Step (3): Listing Procedures or Methodological Techniques

As the last step in RM1, the pronoun *we* served to list procedures or methodological techniques with several particular verbs, as shown in Table 14.

Table 14: Lexical Phrase (1) in RM1-Step (3)

We	<i>next</i> <i>first</i> <i>therefore</i> <i>further</i>	calculated generated injected mapped measured performed quantified selected transfected treated used
<i>Next,</i> <i>First,</i> <i>Therefore,</i>	we	

In addition, the lexical phrase including the adverb *stably* was used to show technical experimental procedures (Table 15).

Table 15: Lexical Phrase (2) in RM1-Step (3)

We	<i>stably</i>	overexpressed knocked down
----	---------------	-------------------------------

3.3 Behavior of Adverbs in RM2

3.3.1 Relationship Between Adverbs and Steps in RM2

RM2 served to announce the results. Observing the concordance lines of 27 adverbs in Table 3 revealed that these adverbs can be divided into two types: 12 adverbs that tended to appear in a sentence-initial position and 15 adverbs that were likely to appear in the middle of sentences. Three steps were identified, which were different from those in Kanoksilapatham (2007). Seven out of 12 adverbs that appeared in a sentence-initial position signaled Step (1) Highlighting important results, while the remaining five adverbs that appeared in a sentence-initial position and the adverb *also* that occurred in a sentence-medial position functioned as Step (2) Showing additional or adversative results. On the other hand, when the 14 adverbs appeared sentence-medially, except for the adverb *also*, they worked as Step (3) Describing quantitative data.

3.3.2 RM2-Step (1): Highlighting Important Results

Table 16 shows seven adverbs, of 12 sentence-initial adverbs, that play an important role in lexical phrases for highlighting important results, and it identifies the number of adverbs that appeared at the initial and middle positions of the sentences in RM2.

Table 16: Adverbs in RM2-Step (1)

Rank	Keyword	LLR	NF	Frequency	Initial position of the sentences (Initial position of RM2)	Middle position of the sentences
1	notably	100.59	1.31	219	199 (49)	20
2	interestingly	100.19	0.60	252	245 (84)	7
3	importantly	62.15	0.67	178	162 (21)	16
4	remarkably	38.76	0.71	67	49 (21)	18
5	strikingly	37.93	0.11	66	58 (27)	8
6	indeed	31.69	1.31	206	157 (93)	49
7	surprisingly	31.12	0.60	62	47 (20)	15

Note: LLR = log-likelihood ratio; NF = normalized frequency.

Although behavior of the adverbs *notably* and *interestingly* were identified with the pronoun *we*, as shown earlier, reading the concordance lines of the other adverbs in Table 16 revealed the same lexical phrase as that presented in Table 17.

Table 17: Lexical Phrase in RM2-Step (1)

<i>Notably,</i> <i>Interestingly,</i> <i>Importantly,</i> <i>Remarkably,</i> <i>Strikingly,</i> <i>Indeed,</i> <i>Surprisingly,</i>	we	found observed	that
	[phenomenon]	was were	found observed

It is worth noting that the lexical phrase navigates readers to what was found in the study, even though the Results sections of medical RAs are considered to be written in an objective way, as Thompson (1993) stated. In addition, when these adverbs sometimes appeared at the initial position of RM2, as shown in Table 16, they functioned not only as signals of highlighting results but also as connectors between RM1 and RM2.

3.3.3 RM2-Step (2): Showing Additional or Adversative Results

Table 18 shows five adverbs of 12 sentence-initial adverbs and the adverb *also* that occurred sentence-medially, which led to lexical phrases for showing additional or adversative results.

Table 18: Adverbs in RM2-Step (2)

Rank	Keyword	LLR	NF	Frequency	Initial position of the sentences (Initial position of RM2)	Middle position of the sentences
1	also	128.86	3.34	1353	19 (2)	1334
2	similarly	71.42	0.53	216	142 (17)	74
3	however	64.14	1.60	649	497 (82)	152
4	furthermore	46.19	0.57	233	228 (28)	5
5	moreover	39.68	0.57	222	217 (23)	5
6	conversely	29.48	0.16	66	53 (10)	13

Note: LLR = log-likelihood ratio; NF = normalized frequency.

The adverbs *similarly*, *moreover*, and *furthermore* were used to form the lexical phrase for adding what was observed (Table 19).

Table 19: Lexical Phrase (1) in RM2-Step (2)

<i>Similarly,</i> <i>Moreover,</i> <i>Furthermore,</i>	we	found observed	that
	[phenomenon]	was were	observed found

Although the adverb *also* appeared in the middle of sentences, as shown in Table 20, it served to form the lexical phrase for adding to the results. Moreover, the adverb *also* with the subject *we* emerged with objective verbs (Table 20).

Table 20: Lexical Phrase (2) in RM2-Step (2)

We	<i>also</i>	detected identified	[substance]		
		noticed	that		
		found	that		
		observed noted	a significant	increase decrease	in

Furthermore, the adverb *also* co-occurred with the nouns *data* and *analysis* and the verb *reveal* (Table 21).

Table 21: Lexical Phrase (3) in RM2-Step (2)

our The This	data analysis	<i>also</i>	revealed	that
--------------------	------------------	-------------	----------	------

Lastly, the adverb *also* was used mainly with the passive voice (Table 22).

Table 22: Lexical Phrase (4) in RM2-Step (2)

was were	also	(significantly)	detected	in by
			evident	
			found	
			induced	
			observed	
			present	
			seen	
			enriched	
			increased	
			decreased	
			reduced	

The lexical phrase in Table 23 shows the relationship between events.

Table 23: Lexical Phrase (5) in RM2-Step (2)

was were	also	associated	with
-------------	------	------------	------

The function of the lexical phrase shown in Table 24 is to connect the results by contrasting what was observed.

Table 24: Lexical Phrase (6) in RM2-Step (2)

<i>However,</i> <i>Conversely,</i>	we	found observed	that
	[phenomenon]	was were	found observed

3.3.4 RM2-Step (3): Describing Quantitative Data

In contrast with the adverbs shown in Tables 16 and 18, except for the adverb *also*, the 14 adverbs in Table 25 tended to appear in the middle of sentences.

Table 25: Adverbs in RM2-Step (3)

Rank	Keyword	LLR	NF	Frequency	Initial position of the sentences (Initial position of RM2)	Middle position of the sentences
1	significantly	664.8	2.53	1026	8 (2)	1018
2	not	382.44	5.43	2198	11 (4)	2187
3	only	125.84	1.78	721	52 (12)	661
4	completely	63.13	0.29	118	1 (1)	117
5	slightly	62.95	0.24	98	0 (0)	98
6	almost	55.65	0.28	115	4 (3)	111
7	markedly	51.03	0.23	93	1 (1)	92
8	alone	48.30	0.48	195	0 (0)	154
9	dramatically	40.12	0.18	76	0 (0)	76
10	nearly	33.03	0.20	81	6 (1)	75
11	consistently	31.85	0.22	91	34 (11)	57
12	substantially	31.36	0.19	80	0 (0)	80
13	approximately	29.90	0.33	134	16 (7)	118
14	strongly	22.65	0.34	141	0 (0)	141

Note: LLR = log-likelihood ratio; NF = normalized frequency.

These sentence-medial adverbs play a critical role in indicating how the event or data quantitatively changed (Table 26). It is worth mentioning that, in terms of the “semantic prosody,” defined by Louw (1993: 157) as “a consistent aura of meaning with which a form is imbued by its collocates,” the adverb *completely* connotated the meaning of “negative” and was associated with the past participle *abolished* or *abrogated*.

Table 26: Lexical Phrase (1) in RM2-Step (3)

was were	(almost) (not)	completely	abolished abrogated blocked depleted eliminated inefficient inhibited lost suppressed	
	(not)	significantly		
		significantly	associated correlated	with
			related	to
		significantly strongly (only) slightly markedly consistently substantially dramatically	higher lower smaller	than in
			decreased increased reduced	in
			more	than

Furthermore, the adverb *approximately* led to the description of data (Table 27).

Table 27: Lexical Phrase (2) in RM2-Step (3)

was were	<i>approximately</i>	two three	times	higher larger smaller	than in
-------------	----------------------	--------------	-------	-----------------------------	------------

The adverb *only* showed limited conditions, such as the location and object (Table 28).

Table 28: Lexical Phrase (3) in RM2-Step (3)

[phenomenon]	was were	significant	<i>only</i>	for
		detected found identified observed increased reduced		in when
	occurred showed			
We	detected found			

The adverbs *almost* and *nearly* were used to show the results in an indirect way (Tables 29 and 30).

Table 29: Lexical Phrase (4) in RM2-Step (3)

was were	<i>almost</i> <i>nearly</i>		identical	
	detected found	<i>almost</i>	exclusively	in

Table 30: Lexical Phrase (5) in RM2-Step (3)

<i>almost</i> <i>nearly</i>	complete	absence	of
--------------------------------	----------	---------	----

Lastly, although Table 6 shows the phrase *alone or in combination*, the phrase usually appeared with the preposition *with* to describe a condition (Table 31).

Table 31: Lexical Phrase (6) in RM2-Step (3)

<i>alone</i>	or	in	combination	with
--------------	----	----	-------------	------

3.4 Behavior of Adverbs in RM3

3.4.1 Relationship Between Adverbs and Steps in RM3

RM3 is used for commenting results. As with RM2, reading concordance lines of 12 adverbs revealed that these adverbs can be divided into two types. Seven adverbs that tended to appear in the initial position of sentences and one adverb that occurred sentence-initially signaled Step (1) “Generalizing/interpreting results,” as stated in Kanoksilapatham (2007: 76). On the other hand, four adverbs that were likely to be found in the middle of sentences worked as Step (2) Emphasizing relationships, which was not mentioned in Kanoksilapatham (2007).

3.4.2 RM3-Step (1): Generalizing/Interpreting Results

Table 32 shows seven adverbs that occurred in the initial positions of sentence and one adverb that appeared in the middle of sentences. These adverbs were strongly related to generalizing the findings.

Table 32: Adverbs in RM3-Step (1)

Rank	Keywords	LLR	NF	Frequency	Initial position of the sentences (Initial position of RM3)	Middle position of the sentences
1	together	960.62	4.69	216	183 (161)	32
2	thus	622.52	4.79	404	324 (287)	80
3	collectively	258.19	1.15	97	83 (78)	14
4	therefore	129.07	1.81	153	108 (90)	45
5	strongly	49.66	0.67	57	0 (0)	57
6	hence	46.25	0.45	38	31 (23)	7
7	overall	39.81	0.66	33	29 (25)	4
8	altogether	38.96	0.18	16	15 (15)	1

Note: LLR = log-likelihood ratio; NF = normalized frequency.

Although the adverbs *together*, *collectively*, and *altogether* with the pronouns *these*, which mark the interpretation, were observed in Table 7, reading the concordance lines revealed similar behavior with the other word combinations, as shown in Table 33. The adverbs shown in Table 32, except for the adverb *strongly*, appeared at the

initial position in RM3. Hence, it can be concluded that not only the verbs *suggest* and *indicate* but also the adverbs shown above were used as connectors between RM2 and RM3. Although the adverb *strongly* appeared only in the middle of sentences, as shown in Table 32, the lexical phrase was sometimes used to emphasize the function of Step (1) with the adverb *strongly* occurring between nouns such as *data* and *results* and verbs such as *suggest* and *indicate* (Table 33).

Table 33: Lexical Phrase (1) in RM3-Step (1)

<i>Together,</i> <i>Collectively,</i> <i>Altogether,</i> <i>Thus,</i> <i>Overall,</i>	these our	data experiments findings observations results	(strongly)	demonstrate indicate show suggest	that
---	--------------	--	------------	--	------

Furthermore, to interpret the data obtained, the adverb *thus* appeared with a participial construction such as *suggesting* and *indicating* (Table 34).

Table 34: Lexical Phrase (2) in RM3-Step (1)

, <i>thus</i>	indicating suggesting	that
---------------	--------------------------	------

In addition, the lexical phrase including the pronoun *we* appeared with the verb *conclude* as shown in Table 35.

Table 35: Lexical Phrase (3) in RM3-Step (1)

We	<i>therefore</i> <i>thus</i>	conclude	that
<i>Therefore,</i> <i>Thus,</i> <i>Overall,</i>	we		

Lastly, the dummy subject *it* played a critical role in adjusting the interpretation with the adverbs *therefore*, *thus*, and *hence* (Table 36).

Table 36: Lexical Phrase (4) in RM3-Step (1)

<i>Therefore,</i> <i>Thus,</i> <i>Hence,</i>	it	appears seems		that
		is	likely possible	

3.4.3 RM3-Step (2): Emphasizing Relationships

Table 37 shows the adverbs that were present in the middle of sentences as keywords of RM3, except for the adverb *strongly* (Tables 32 and 33). Although the adverb *rather* was found with the conjunction *than*, the phrase *rather than* was not defined as a lexical phrase consisting of more than two words.

Table 37: Adverbs in RM3-Step (2)

Rank	Keywords	LLR	NF	Frequency	Initial position of the sentences (Initial position of RM3)	Middle position of the sentences
1	not	122.97	5.73	483	0 (0)	483
2	rather	64.97	0.78	66	2 (2)	64
3	directly	27.21	0.78	66	0 (0)	66
4	functionally	22.99	0.27	23	0 (0)	23

Note: LLR = log-likelihood ratio; NF = normalized frequency.

The lexical phrase in Table 38 included the adverb *directly*, which indicated physical interaction or action without intervening events.

Table 38: Lexical Phrase (1) in RM3-Step (2)

bind	<i>directly</i>	to
interact		with

Furthermore, with the adverb *functionally*, the lexical phrase emphasized the characteristics of the research object as a technical term (Table 39).

Table 39: Lexical Phrase (2) in RM3-Step (2)

is are	<i>functionally</i>	essential required	for
		distinct	in

4. Discussion and Conclusion

Using a corpus based on move analysis, we investigated the behavior of adverbs in the Results sections of experimental medical RAs. We produced 4-grams with adverbs and observed the concordance lines of adverbs carefully while referring to the LSD. Accordingly, we identified 26 lexical phrases for realizing three steps in RM1, three steps in RM2, and two steps in RM3. In these lexical phrases, the adverbs occurring in the initial positions of sentences sometimes functioned as connectors between moves, for example, the adverbs *interestingly* in RM2 and *collectively* in RM3. On the other hand, adverbs in the middle of sentences such as *significantly* were used to show the quantitative data in RM2.

However, our analysis of adverbs has several limitations. For instance, we did not observe the concordance lines of the adverbial phrases *as expected* and *of note* in RM2 and the adverbial phrase *taken together* in RM3. The former behaved in a similar manner as that of the adverb *interestingly*, following the phrase *we found*, to highlight important results (Table 17). The latter phrase seemed to function similarly to the adverb *therefore*, following the phrase *we conclude*, to generalize the results (Table 35). We may be able to find other lexical phrases by exploring the behavior of adverbial phrases or other parts of speech in RAs. Moreover, future work should consider how the frequencies and example sentences of lexical phrases are shown.

Lastly, our results provide new insights into teaching the art of writing RAs. The most influential element in teaching English is creating vocabulary lists (Groom & Littlemore, 2011), such as a New Academic Word List (Coxhead, 2000), a New Academic Vocabulary List (Gardner & Davies, 2014), and a New Medical Academic Word List (Lei & Liu, 2016). In addition, Simpson-Vlach and Ellis (2010) and Martinz and Schmitt (2012) presented lists of pedagogically useful expressions for academic English as an Academic Formulas List and a Phrasal Expressions List, respectively. However, these lists do not indicate the behavior of words or the relationships between

the lexical phrases and the functions of steps. As Le and Harrington (2015) mentioned, learners need to recognize frequent word combinations in specific moves or steps in a particular discourse community. Although further studies are needed to determine whether the lexical phrases obtained are observable for other disciplines and to measure their effectiveness, the lexical phrases described here according to the functions of steps should be of use for teaching English for specific purposes.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K23095.

Notes

1. Available from <http://langtest.jp/awsum/>
2. Although four articles contained Results and Discussion sections, we identified Results moves and Discussion moves appropriately.
3. Available from <https://lsd-project.jp/cgi-bin/lsdproj/ejlookup04.pl>.
4. PubMed is a web site that contains 30 million citations of Abstracts related to biomedical literature (available from <https://www.ncbi.nlm.nih.gov/pubmed/>).

Appendix

The following is a list of the 30 selected journals with the number of articles from each in parentheses.

Brain (11), Cancer Cell (10), Cancer Research (10), Cell (12), Cell Metabolism (10), Cell Reports (10), Cell Stem Cell (10), Cellular Microbiology (11), Current Biology (11), Developmental Cell (10), EMBO Journal (11), Genes & Development (11), Genome Research (10), Journal of Biological Chemistry (8), Journal of Cell Biology (11), Journal of Experimental Medicine (10), Journal of Neuroscience (11), Journal of Virology (11), Molecular and Cellular Biology (10), Molecular Cell (11), Nature Cell Biology (10), Nature Genetics (9), Nature Immunology (10), Nature Medicine (9), Nature Neuroscience (10), Nature Structural & Molecular Biology (10), Neuron (10), Oncogene (10), PLOS Biology (8), and PLOS Genetics (9)

References

- Anthony, L. (2015) TagAnt (Version 1.2.0) [Computer Software]. URL: <https://www.laurence-anthony.net/software/antconc/>
- Anthony, L. (2019) AntConc (Version 3.5.8) [Computer Software]. URL: <https://www.laurenceanthony.net/software/antconc/>
- Biber, D. (1990) "Methodological Issues Regarding Corpus-Based Analyses of Linguistics

- Variation." *Literary and Linguistic computing* 5, 4: 257-269.
- Biber, D. (1993) "Representativeness in Corpus Design." *Literary and Linguistic computing* 8, 4: 243-257.
- Biber, D., S., Conrad, and Cortes, V. (2004) "If You Look at ...: Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics* 25, 3: 371-405.
- Biber, D., U. Connor, and Upton, T. A. (2007) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins Publishing Company.
- Brett, P. (1994) "A Genre Analysis of the Results Section of Sociology Articles." *English for Specific Purposes* 13, 1: 47-59.
- Case, E. D., Chong, A., Wehrly, T. D., Hansen, B., Child, R., Hwang, S., Virgin. H. W., and Celli, J. (2014) "The Francisella O-antigen Mediates Survival in the Macrophage Cytosol via Autophagy Avoidance." *Cellular Microbiology* 16, 6: 862-877.
- Cheng, W. (2012) *Exploring Corpus Linguistics Language in Action*. New York: Routledge.
- Coffin, C. (2001) "Theoretical Approaches to Written Language - A Tesol Perspective." In Burns, A., & Coffin, C. (eds.), *Analysing English in a Global Context*. New York: Routledge, pp.93-122.
- Coxhead, A. (2000) "A New Academic Word List." *TESOL Quarterly* 34, 2: 213- 238.
- Dudley-Evans, T., and St John, M. J. (1998) *Developments in English for Specific Purposes: A Multi-Disciplinary Approach*. Cambridge University Press.
- Durrant, P. (2009) "Investigating the Viability of a Collocation List for Students of English for Academic Purposes." *English for Specific Purposes* 28, 3: 157-69.
- Gardner, D., and Davies, M. (2014) "A New Academic Vocabulary List." *Applied Linguistics* 35, 3: 305-327.
- Gledhill, C. (2000a) "The Discourse Function of Collocation in Research Article Introductions." *English for Specific Purposes* 19, 2: 115-35.
- Gledhill, C. (2000b) *Collocations in Science Writing*. Verarbeitung: Gunter Narr Verlag Tübingen.
- Gray, V., and Biber, D. (2015) "Phraseology." In Biber, D., & Reppen, R. (eds.), *The Cambridge Handbook of English Corpus Linguistics*: Cambridge University, pp.125-145.
- Groom, N., and Littlemore, J. (2011) *Doing Applied Linguistics: A Guide for Students*. New York: Routledge.
- Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge University press.
- Hyland, K. (2002) "Specificity Revisited: How Far Should We Go Now?" *English for Specific Purposes* 21, 4: 385-395.
- Hyland, K. (2008a) "As can Be Seen: Lexical Bundles and Disciplinary Variation." *English for Specific Purposes* 27, 1: 4-21.
- Hyland, K. (2008b) "Academic Clusters: Text Patterning in Published and Postgraduate Writing." *International Journal of Applied Linguistics* 18, 1: 41-62.
- Imao, Y. (2019) CasualConc (Version 2.1.2) [Computer Software]. URL: <https://sites.google.com/site/casualconc/download>
- Kanoksilapatham, B. (2005) "Rhetorical Structure of Biochemistry Research Articles."

- English for Specific Purposes* 24, 3: 269–92.
- Kanoksilapatham, B. (2007) “Rhetorical Moves in Biochemistry Research Articles.” In Biber, D., Connor, U., & Upton, T.A. (eds.). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins, pp. 73–119.
- Kawamoto, T., and Ishii, T. (2018a) *Top-jyanajyu-395hen-no-kata-de-kaku igakueigoronbunn*. Tokyo: Yodosha.
- Kawamoto, T., and Ishii, T. (2018b) “Move Analysis of English Medical Papers and Its Application to the Writing of the Introduction and Discussion Sections.” *Journal of Medical English Education* 17, 3: 107–111.
- Le, T. N. P., and Harrington, M. (2015). “Phraseology Used to Comment on Results in the Discussion Section of Applied Linguistics Quantitative Research Articles.” *English for Specific Purposes* 39, 45–61.
- Lei, L., and Liu, D. (2016) “A New Medical Academic Word List: A Corpus-Based Study with Enhanced Methodology.” *Journal of English for Academic Purposes* 22: 42–53.
- Louw, B. (1993) “Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies.” In Bake, M, Francis, G., & Tognini-Bonelli, E. (eds.), *Text and Technology: In Honor of John Sinclair*. Amsterdam: John Benjamins Publishing Company, pp.157–175.
- Malcolm, L. (1987) “What Rules Govern Tense Usage in Scientific Articles?” *English for Specific Purposes* 6, 1: 31–47.
- Martinz, R., and Schmitt, N. (2012) “A Phrasal Expressions List.” *Applied Linguistics* 33, 3: 299–320.
- Mizumoto, A., Hamatani, S., and Imao, Y. (2016) “Analysis of Research Articles in Applied Linguistics Using an Approach Combining Moves and Lexical Bundles: Towards Developing a Support Tool for Writing Research Articles in English.” *English Corpus Studies* 23: 21–32.
- Nwogu, K. N. (1997) “The Medical Research Paper: Structure and Functions.” *English for Specific Purpose* 16, 2: 119–138.
- Saber, A. (2012) “Phraseological Patterns in a Large Corpus of Biomedical Articles.” In Boulton, A., Carter-Thomas, S., & Rowley-Jolivet, E. (eds.), *Corpus-Informed Research and Learning in ESP: Issues and Applications*. Amsterdam: John Benjamins Publishing Company, pp.45–81
- Simpson-Vlach, P., and Ellis, N. (2010) “An Academic Formulas List: New Methods in Phraseology Research.” *Applied Linguistics* 31, 4: 487–512.
- Sinclair, J. M. (2004) *Trust the text*. London: Routledge.
- Swales, J. M. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Tasdemir-Yilmaz, O. E., and Freeman, M. R. (2014) “Astrocytes Engage Unique Molecular Programs to Engulf Pruned Neuronal Debris from Distinct Subsets of Neurons.” *Genes & Development* 28, 1: 20–33.

- Thomas, S., and Thomas, P. H. (1994) "Reporting Verbs in Medical Journal Articles." *English for Specific Purposes* 13, 2: 129–48.
- Thompson, D, K. (1993) "Arguing for Experimental "Facts" in Science." *Written Communication* 10, 1: 106–124.
- Waard, A., and Henk, P. H. (2012) "Verb Form Indicates Discourse Segment Type in Biological Research Papers: Experimental Evidence." *Journal of English for Academic Purposes* 11, 4: 357–66.
- Williams, I, A. (1996) "A Contextual Study of Lexical Verbs in Two Types of Medical Research Report: Clinical and Experimental." *English for Specific Purposes* 15, 3: 175–197.
- Williams, I, A. (1999) "Results Sections of Medical Research Articles: Analysis of Rhetorical Categories for Pedagogical Purposes." *English for Specific Purposes* 18, 4: 347–366.

「論文」

アメリカ英語における *be bound to* の準助動詞化の過程*

家口美智子

Abstract

This study clarifies the process of grammaticalization of *be bound to* from a lexical phrase to a semi-auxiliary through the analysis of the Corpus of Historical American English. It is found that its epistemic usage is considered to be a full-fledged semi-auxiliary in present-day English, whilst its obligation usage still expresses a lexical function in certain tokens. The grammaticalization started with the increased use of inanimate subjects in the 1840s, followed by the emergence of sentences forming passive voice in the 1870s and *there* existential sentences in the 1880s. Then the structures of subjects containing *what*, such as *What happened yesterday is bound to be a lucky sign* and *What is bound to happen will be a lucky sign* appeared around 1900, earlier than gerund subjects (e.g. *Leading a good life was bound to be difficult*), which began to be used in the 1940s and *to* infinitive subjects (e.g. *To lead a good life was bound to be difficult*), which started in the 1960s. Judging from the frequent use of *there be bound to* and *be bound to happen* and the declining tendency of *be bound to* since the 1950s, this study concludes that it is and will remain in the phase of ‘static’ grammaticalization in contemporary American English.

1. はじめに

be bound to は義務を表すときと認識的な意味を表すときとで文法的なふるまいが異なる面白いフレーズである。Quirk et al. (1985: 137) は *be bound to* を準助動詞 (semi-auxiliary) として分類しているが、¹ 往々にして先行研究は準助動詞の一つとして簡単な説明に終始している。当該フレーズを後期近代英語以降の通時的な変化から検証すると、準助動詞化する過程での興味深い現象を明らかにすることができる。本論は、*be bound to* の詳細な記述を行うとともに、歴史的発達を提示し、他の準助動詞の発達の解明の一助となることを目指す。

まず, be bound to の辞書における記述を見ておく。*Longman Dictionary of Contemporary English* 6 版 (以下 LDOCE⁶) は, bound を形容詞として見出し語で扱い, 第 1 の意味として be bound to というフレーズの形で, ‘to be very likely to do or feel a particular thing’ であると定義し, ‘Don’t lie to her. She’s bound to find out.’ という例をあげている。この認識的用法は, 話し言葉で主に使われるとしている。第 2 の意味として, be bound (by sth) という形を基本として提示し, ‘to be forced to do what a law or agreement says you must do’ と定義し to 不定詞と共に共起する例で ‘The foundation is bound by the treaty to help any nation that requests aid.’ をあげている。第 3 の意味として, be/feel bound to do something の形で, ‘to feel that you ought to do something because it is morally right or your duty to do it’ と説明し, ‘Well, I’m bound to say, I think you’re taking a huge risk.’ という例や be duty/honour bound to do sth という定型表現をあげている。第 2 と第 3 は義務を表す用法である。

アメリカ英語でも同じような意味で機能している。*Longman Dictionary of American English* 5 版 (以下 LDOAE⁵) は形容詞 bound の見出し語として, be bound to の形で, ‘to be certain to do something’ という認識的な用法を第 1 の意味として記述し, 第 2 の意味として義務的な用法で be bound to という形を明示することなく, ‘having a legal or moral duty to do something’ として ‘The company is bound by law to provide safety equipment.’ などの例をあげている。

Oxford English Dictionary 2 版 (以下 OED) は過去分詞及び形容詞の見出し語の bound は 2 つあるが, 2 つめ (s.v. bound: ppl, a²) の第 7 の意味で, 義務を表す場合のカテゴリーのもと, 不定詞と共に共起する be bound to を ‘compelled, obliged; under necessity (esp. logical or moral); fated, certain; also in U.S. determined, resolved’ という定義を提示し, 義務的な用法と認識的な用法があるとしている。OED は義務的用法と認識的用法で項目を分けていない。OED の説明によると方言によっては bound の代わりに tied が使用されることもあるとあり, 「縛る」という意味からこの義務的用法が発達したと思われる。この見出し語の中では 14 世紀の例が初出である。また, bound の 1 つめ (s.v. bound: ppl, a¹) の第 3 の意味の説明で, 電車等の行き先を表すのに使われる bound for というフレーズのもとになった boun という動詞を基としてできた bound が to 不定詞と共に共起して ‘about to’ ‘going to’ という意味で方言で使われるが, 本論で議論している動詞 bind からできた認識的用法の be bound to はこのフレーズから影響を受けたのではないかと OED は推論している。1862 年と 1864 年の例を 2 つあげて

いる。

義務的用法と認識的用法の頻度に関して、Palmer (1990: 55) が be bound to のほとんどは認識的な用法で使用されると説明している。Mong (n.d.) は、Cambridge Dictionaries Online を 2012 年に分析し be bound to の 5,725 例のデータから無作為に抽出した 100 例の内、92 例が認識的、5 例が義務的用法で使われていると報告している。また、Mong (n.d.) のデータでは LDOCE⁶ や LDOAE⁵ の説明や例文にあるように、by 句と共に起して義務を課すものが表現されているものが 48 例あったということだ。しかしながら実際は、限られた文脈から 2 つの意味を正確に判別することは時に難しい。例えば、Westney (1995: 149-150) は be bound to の意味を説明する際に、両者のどちらか判別ができないケースがあることを複数例示している。

be bound to は後期近代英語から現代英語において大きく変化を起こしているため、この時期の 4 億語という大量のデータを収めている Corpus of Historical American English (以下 COHA) を本論は分析し 1810 年代以降現在に至るまでの、be bound to が準助動詞化する過程を明らかにする。また、文法的なふるまいを詳細に記述する。他のコーパスでは COHA に比する同時期の通時的な大量データを提示していないため、本論は COHA に特化し、アメリカ英語における be bound to の文法化に焦点を当てる。本論は、Westney (1995: 11) にならい、「準助動詞」を、1) grammaticalization, 2) idiomaticity, 3) semantic relatedness to a central modal auxiliary の性質を持つフレーズであると定義する。詳しくは 2 節で紹介する。また、「文法化」を Hopper and Traugott (2003: iv) が提唱する、‘the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions’、すなわち語彙項目や構文がある文脈で文法的な機能を担うようになり、それが一旦文法化されると、より新しい機能が発達する過程であると定義する。また、「準助動詞化」は主には文法化によって牽引されるという立場で議論を行う。本論は、「認識」、「義務」を話者の主観性・客観性に関係のない広義の意味として用いる。また、「フレーズ」は、一定のまとまりのある 2 つ以上の語の連鎖を幅広く呼ぶ。「チャンク」とはフレーズより更に語の結びつきが強くなり、構成要素が分析されることなく、分解できない一つの塊として機能し、全体として処理される 2 つ以上の語の連鎖を意味するものとする。

本論は、Quirk et al. (1985: 137) が同じく準助動詞として分類をしている be obliged to と比較をすることで be bound to の準助動詞化の過程の特徴を際立た

せる方法を取る。2節で先行研究を概観した上で更に詳細な記述を試み、3節で文法化のパターンを確認し、4節で調査結果を提示し考察を行い、5節で結ぶ。

2. 現代英語における be bound to : be obliged to との比較より

まず be bound to についての先行研究を概観する前に、be obliged to というフレーズについても辞書の記述を簡単に見ておく。LDOCE⁶で、他動詞としての見出し語 oblige のもと、‘[usually passive] if you are obliged to do something, you have to do it because the situation, the law, a duty etc makes it necessary’ という説明があり、通常は受動態の形で義務を表すとしている。‘The minister was obliged to report at least once every six months.’ などの例があげられ、日常会話では通常使われない formal な語であると説明されている。be obliged to という形が明記されているわけではない。一方、LDOAE⁵は、見出し語の動詞 oblige のもと、第1の意味として、be/feel obliged (to do sth) の形を紹介して、‘to feel that it is your duty to do something’ と説明している。OED も同様に見出し語の動詞 oblige の第9番目に他動詞 oblige が受動態になり to 不定詞と結びついた例を紹介しているが、能動態の例と同じに取り扱われている。受動態としては1776年の例がこの見出し語の例では初出である。

もともと「compel : 強いる」という意味の他動詞 bind と oblige が受動態になることで、主語に義務を課すことを表す文型となったことがわかる。Westney (1995: 1-37) も両フレーズを must が原義であると説明している。ところが、LDOCE⁶、LDOAE⁵やOEDの定義で見たように、現代英語では be bound to は認識的な蓋然性を表現することが主な機能であるのと対照的に、be obliged to は義務を表す意味のみである。be bound to が認識的な用法を発達させたのは、Bybee (2015: 122-124) が主張する未来 (future (prediction)) を表す grammatical morphemes は義務を表す語彙的なものから発達するという文法化の一方向性 (unidirectionality) に沿う代表的な例であろう。ところが、現代英語では be obliged to は be bound to と異なり認識的用法まで発達を遂げていないと言える。

Westney (1995: 1-37) は Quirk et al. (1985: 137) が be bound to と be obliged to を同じ準助動詞に分類していることに異議を唱え、以下の6つの理由により be bound to のみが準助動詞であると議論している。以下の(1)から(6)の例文は、Westney (1995) から採用した以外の例は、Corpus of Contemporary American English (以下 COCA) から採用したものと、出典の記述がないもの

は Westney (1995) の説明に基づいて筆者が例を作成し、文法性をインフォーマント（アメリカ人2名、イギリス人2名）に確認したものである。

第一に、準助動詞としての性質があると、主語の選択に制限がなくなる。

- (1) a An influx of new ideas is bound to be disruptive. (COCA, fiction)
- b It's bound to rain sooner or later. (COCA, fiction)
- c *An influx of new ideas is obliged to be disruptive.
- d *It's obliged to rain sooner or later.

(1a) のように、be bound to は概念的な無生物主語 (inanimate subject) が使用可能であるし、(1b) が示すように、天候を表す it も取ることができる。ところが、(1c) や (1d) にあるように be obliged to は意志を持つ性質を持つもの (animate subject) しか通常は主語になれない。

第二に、準助動詞としての性質があると、there 存在文で使用可能となる。よって be bound to は存在文の繰り上げ構文として使用されるが、be obliged to は there 存在文では使用されない。

- (2) a The classes are huge which is good because there is bound to be one person that can help you. (COCA, academic)
- b ??There is obliged to be a teacher for every 20 students.

第三に、準助動詞としての性質があると、態が変わっても基本的な意味は変わらない。

- (3) a The buildup in the supply dump was bound to be noticed. (COCA, fiction)
- a' They were bound to notice the buildup in the supply dump.
- b But he just stood there, so she was obliged to look up at him again. (COCA, magazine)
- b' But he just stood there, so he was obliged to be looked up at by her again.

(3a) と (3a') では主語が交代しても文が表す基本的な意味は変わらないが、(3b) と (3b') では、能動態と受動態では表現される意味が異なる。

第四に、準助動詞としての性質があると、being を従えない場合は文頭で使

われない。(4a, b) のように文頭で bound to は容認度が若干下がるが obliged to は文法的となる。

- (4) a ?Bound to take stern measures, the administration lost popularity.
(Westney, 1995: 19)
- b Obligated to leave Beirut when the Palestinians were compelled to depart in 1982, she wonders what she can write about the experience of being in exile.
(COCA, academic)

第五に、準助動詞としての性質があると、最後の to が抜けると容認度は低い。

- (5) a ??This game is bound to be cancelled because of the bad weather, but that game is not bound.
- b Her leave is recommended, but she is not obliged.

(5) のように、be bound to は to がないと非文になることから be bound to でチャックとして存在しているといえる。

第六に、準助動詞としての性質があると音韻上の削除 (phonological reduction) が起きる。be bound to は [d] の脱落 (elision) が起き bound と to が融合同化 (coalescent assimilation) して発音されうるが、be obliged to はそうではない。

- (6) a *be bound to* ('certain') – cf. *be found to*
- | | |
|-------------------------|----------------------------|
| he's bound to come soon | this was found to be false |
| ... /baʊndtə/... | ... /faʊndtə/... |
| ... /baʊntə/... | ... */faʊntə/... |
- (Westney 1995: 34)
- b he was obliged to come
- | | |
|----------------------|--|
| ... /ə'blaɪdʒdtə/... | (LDOCE ⁶ と LDOAE ⁵ より) |
| ... */ə'blaɪtə/... | |

以上が Westney (1995: 1-37) があげた 2 つのフレーズに関した差である。² be bound to は意味の側面では、主語になる動作主に課せられた義務を表す本来の用法が文法化を経て「確かに<・・する>はず (certain) to do」(新英和大辞

典 6 版) という認識的な蓋然性を表すようになり、同時に文法的なふるまいも変化していると言える。Hopper & Traugott (2003: 127) が説明する文法化の特徴である ‘semantic fading, phonological reduction, positional fixing, erasure of word boundaries’ という 4 つの現象を全て満たしている。一方, be obliged to は oblige の動詞としての意味に基づいた文法的・統語的なふるまいをしている。

ここで本論は be bound to の準助動詞化あるいは文法化を考える時に義務的用法と認識的用法の差として留意する点をおさえておきたい。Westney (1995: 1-37) は上で見た (1) から (6) の特徴に関して認識的用法と義務的用法の差について明確にしないまま議論しているので、もう少し丁寧に議論する必要がある。実は彼も認識的用法と義務的用法では文法的なふるまいが異なることを暗に指摘している。(7) から (9) は全て Westney (1995: 21) からの例である。

- (7) a He's bound to give the right answer.
 b *He's bound by his knowledge to give the right answer. ('he's certain')
- (8) a If you believe this, you're bound to act on it.
 b If you believe this, you're bound by your conscience to act on it.
 ('you're obliged')
- (9) a He was obliged to settle for early retirement.
 b He was obliged by his employers to settle for early retirement.

(以上, Westney (1995: 21))

Westney (1995: 21) は (7) のように認識的用法を表す場合は, be と bound の間に by 句が入ると (7b) は非文となるのに, (8) のように義務的用法を表す場合は, (8b) のように文法的であると記述しているが, つまりは, 認識的な意味を表現するときは be bound to はチャンクとしてふるまうと主張しているのに, 義務を表す場合はチャンクとして機能しているのでなく, bind という語を中心にできた受動態の文であると語らずして示している。また, (6a) の音韻論的な削除の現象でも彼は ‘certain’ という認識的な意味に制限している。これらの事実に加え, 本論はチャンク性に関して, 更に 3 つの点を提示したい。まず, be bound to の be と bound の間に not 以外の副詞が 1 個現れる場合を考える。³ COCA で 141 例あるが, LDOCE⁶ の記述で見たように, be duty bound to (28 例), be honor bound to (11 例) の「道徳的に / 義務的に・・する必要がある」という意味の定型表現が頻繁に使われ, 次いで legally (18 例), morally (12 例),

contractually (6 例), constitutionally (1 例) というように計 80 例が明らかに義務を表す例となっている。残りの 61 例は also, all, always, therefore, indeed などが挿入されている。義務の用法と認識的用法のどちらであるかに関して正確な判断が難しいので詳しい分析は避けるが、たとえこれらの 61 例全てが認識的用法で使用されているとしても、Palmer (1990: 55) に従えば、もともと義務的用法は割合が少ないことを考えると、半数以上 (56.7%) が義務を表している事実は、義務的用法の *be bound to* がチャンクとして確立していないことを示している。2 点目は、先ほど見た (4) の例で Westney は区別をしていないが、分詞の部分が義務を表すときには、インフォーマントによると (10) のように *being* がなくても文法的に正しい文となる。(10) は筆者の作例である。

- (10) Bound to support a ban on abortion, Tom, a Catholic, didn't vote for the Democratic candidate, even though he liked all their economic policies.

3 点目として、(5) の例で見た *to* の脱落である。認識的な意味の時には *be bound to* はチャンクとして機能するが、義務を表す場合はそうではないケースが (11) のようにコーパスで散見されるし、インフォーマントも文法的であるとする。

- (11) Weicker will not stand still for such questions. He will not be bound. But by November, the voters may have decided that they are not bound to cast their votes for Roger Eddy's staunch supporter. (COCA, magazine)

これらのことから、他動詞 *bind* の語としての意味が表現される義務を表す場合と認識を表す場合では Westney (1995) のあげた 6 つの基準での準助動詞である度合いが異なると言える。ただし、義務を表現する場合でも、(2) のような *there* 存在文での用例はあるし、(3) のような態の変化で基本的な意味は変わらない解釈が可能である。

3. 文法化のパターンについて

本節では、OED の 2 版を分析した Mair (2004) に基づいて、文法化には、dynamic な文法化と static な文法化があることを見る。Mair (2004) は dynamic な例として、be going to, start (アスペクトを表現する準助動詞として), help (to) 動詞, static な例として、分詞構文で使用される seeing (that), supposing (that) をあげている。本論では be going to と seeing (that) を見る。図 1 のパターンで見られるように、be going to は、Jespersen (1932: 217) の調査によると文法化は 15 世紀末に始まったが、1600 年頃でもまだ頻度は低かった。1800 年代後半から急激に頻度が増え、1900 年代に入って gonna も使用されるようになっていく。一方、static な文法化を示す seeing (that) は、中英語後期にすでに文法化をしていたが、ここ 500 年はさほど頻度は増えていない状態である(図 2 の一番下の実線)。Mair (2004) は非常に頻繁に使用される動詞 see のもと、seeing (that) は文法化をしているものの頻度を増やすことなくひっそりと英語で生き残っていると説明している。

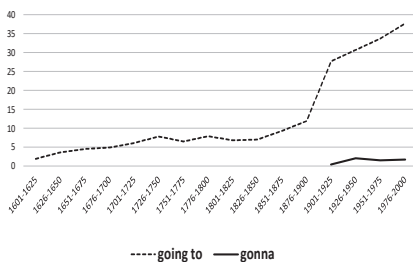


図 1 *Going to and gonna* 1600-2000

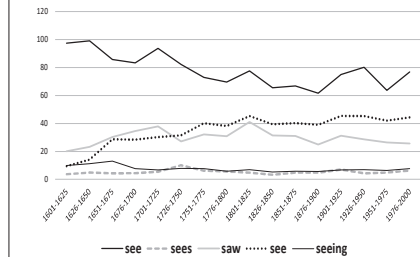


図 2 *See* 1600-2000

頻度：OED の quotation text 1 万例あたりの出現回数

図 1 は Mair (2004: 129), 図 2 は Mair (2004: 134) より抜粋

文法化は英語史において早い時期から始まっているということ、また 2 種類の文法化のパターンがあることを念頭において次節から be bound to が準助動詞へと変貌する過程を見ていく。

4. be bound to の文法化の過程

本節では be bound to と be obliged to の 1810 年以降の文法化について COHA を分析して文法化を検証する。まずは頻度を見る。be bound/obliged to+ 動詞を抽出するのに、be bound to は {[vb*] bound to [v?i*]}, be obliged to は、 {[vb*] obliged to [v?i*]} で検索をかけている（レマ表示）。be 動詞の縮約形’s もこの検索式で取り込めている。全てのデータを手作業で見て判断を行い、動詞の代わりに名詞が使用されているものは削除した。図 3 は be bound to と be obliged to の頻度をグラフで表したもので、表 1 は 2 つの生起数と割合である。

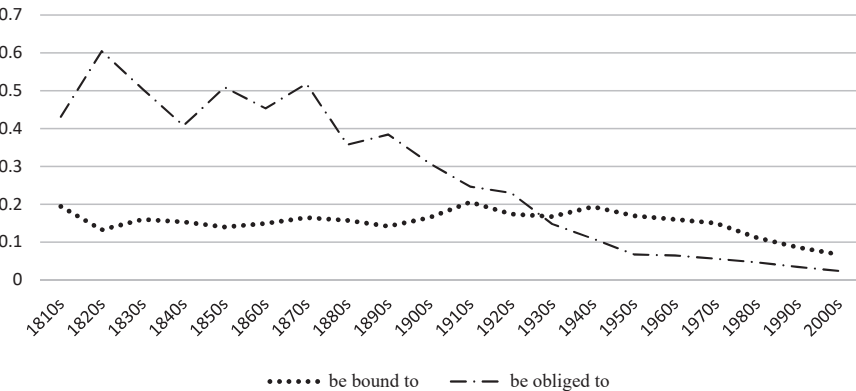


図 3 be bound to と be obliged to の1810年以降の推移

表 1 be bound to と be obliged to の1810年以降の生起数と頻度（1 万語あたり）

	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
be bound to 生起数	23	92	221	247	232	257	307	330	302	373	467	448	410	468	414	384	358	283	241	200
be obliged to 生起数	51	419	695	655	840	777	966	747	815	697	560	591	362	265	166	156	134	119	97	72
be bound to 頻 度	0.19	0.13	0.16	0.15	0.14	0.15	0.16	0.16	0.14	0.17	0.21	0.17	0.17	0.19	0.17	0.16	0.15	0.11	0.09	0.07
be obliged to 頻 度	0.43	0.60	0.50	0.41	0.51	0.45	0.52	0.36	0.38	0.31	0.25	0.23	0.15	0.11	0.07	0.07	0.06	0.05	0.03	0.02

COHA に収められた 200 年のデータでは be bound to が 6,057 例、be obliged to が 9,184 例あり、後者の方が頻度が高い。ところが、COCA における書き言葉

では be bound to のほうが約 3 倍頻繁に現代英語では使われている。(COCA では 3,307 例, be obliged to は 1,163 例である。spoken を除いた fiction, magazine, newspaper, academic での頻度は, それぞれ 2,861 例, 1,068 例と前者のほうが 2.5 倍以上使用されている。⁴⁾ 上のグラフで, COHA の 2000 年代のデータに注目すると, 頻度が前者は 0.067 回 /1 万語, 後者は 0.024 回 /1 万語で, 現代英語では be bound to は約 3 倍の頻度を示していて, COCA のデータとの矛盾はない。図 3 で明らかのように, 1920 年代くらいまで be obliged to のほうが頻度が高かったため, 総計数では逆の数値が出てきたと言える。

be bound to は 1900 年代頃から頻度が漸増し, 1940 年代くらいにピークを迎えるが, その後なだらかに減少し続けている。一方, be obliged to は 1870 年代あたりまで頻繁に使われ 1890 年代以降使用される頻度が下がっている。ともに, 現代のアメリカ英語では使用が減りつつあるフレーズであると言える。

4.1 be bound to の文法化

be bound to の分析結果を見てみる。まず, 主語の性質についてである。まず, 無生物主語の割合を分析する。分析方法を簡単に紹介する。justice, economy のような抽象名詞は明らかに無生物主語であるが, 難しい例がある。例えば, city は be bound to の不定詞に使われる動詞の意味により判断が異なる。

(12) a ...the city will not be bound to turn over its properties to the new agency...

(COHA, news, 1953 年)

b Every large city in the world was bound to have a characteristic street
or square.

(COHA, fiction, 1957 年)

(12a) は city が「土地などを新しい部局に明け渡す」という意味でメトニミーとして意思を持っている例であるとみなし生物性を持つ主語であると判断した。一方, (12b) では city が意思をもたず客観的な記述がされていると判断できるので無生物主語としてカウントした。bank, committee, company, congress, council, court, government, judiciary, legislature, state などほぼメトニミーとして使用されているため, 無生物主語ではなく意思がある生物主語としてはほぼカウントした。一方, city 同様に, 国名や市の名前, army, newspaper, school, university など文脈ごとの動詞の意味で両者に振り分けた。無生物主語かどうか文脈が不十分であるため判断がつかなかったものは分析から除外し

ている。⁵ 無生物主語かつ抽象名詞であるものの COHA の初出を以下にあげておく。

- (13) In other words, the lowest form of existence, mere Life, is bound to preserve its elevation, and not communicate with Death.⁶

(COHA, magazine, 1848 年)

図 4 は無生物主語が be bound to の総数に占める割合のグラフである。表 2 は生起数と割合のデータを示している。

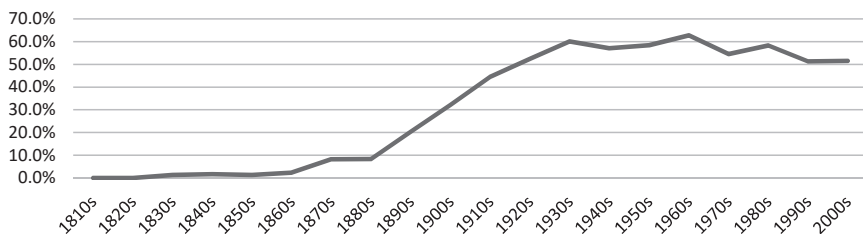


図 4 無生物主語の割合

表 2 be bound to の総数, 無生物主語の生起数, 無生物主語の割合

	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
be bound to の総数	22	91	220	243	232	254	304	278	300	370	465	446	404	462	411	381	356	278	240	200
無生物主語の生起数	0	0	3	4	3	6	25	23	61	119	207	234	243	264	240	239	194	162	123	103
無生物主語の割合	0.0%	0.0%	1.4%	1.6%	1.3%	2.4%	8.2%	8.3%	20.3%	32.2%	44.5%	52.5%	60.1%	57.1%	58.4%	62.7%	54.5%	58.3%	51.3%	51.5%

上のグラフを見ると、1870 年代頃から割合は増え始め、1910 年代頃より半数は無生物主語が使われている。(1) で見たように主語の選択に制限がなくなっていく様子がうかがえる。前述のとおり、認識的用法と義務的用法の正確な区別は時に難しいので正確な分析はできない。しかしながら、無生物主語に義務を課すよりは生物主語に課すほうが社会通念上一般的で、特に抽象的な名詞(例: equality, justice, method, departure)が主語になった場合は受動態の文で現れていない限り義務的用法であると解釈することは時に難しい。よって、無生物主語が増えることは、認識的用法が増えている傍証になる。

統語的な特徴も顕著である。まず、1870 年代に受動態が使用されるようになってくる。

- (14) a But they are bound to be guided entirely by their constitutional advisers...
 (COHA, magazine, 1870 年)
 b ...yet justice, by the constitution of England, is bound to be administered
 in mercy... (COHA, magazine, 1875 年)

(14a) にあるように 1870 年に生物主語である ‘they’ (governors) が受動態での主語として使用され始めたが、1875 年では (14b) にあるように無生物主語の justice が受動態で使用されている。但し、(14b) の例は義務的用法であろう。be bound to 自体が受動態であることから、受動態が重なるのはスタイル的には理想的ではないので、当該フレーズがチャンクとして機能している傍証となる。

次に (2) で見た there 構文での使用を見る。COHA において初出の例は 1881 年で 2 例ある。その内の 1 例は以下のとおりである。

- (15) ...and there’s bound to be plenty of sickness there, sooner or later.
 (COHA, fiction, 1881 年)

4.3 でも言及するが、there 構文での使用はこれから飛躍的に増えていく。be 動詞だけでなく、他の動詞も (16) のように 6 例使われている。

- (16) There is bound to come a cropper somewhere. (COHA, fiction, 1920 年)

次に主語の抽象性について調べる。1890 年代より、(17) のように what 節が主語になる be bound to が使われている文が現れている。

- (17) And whatever she would say was bound to have a quality of interest and attraction that could be exercised by no other lips. (COHA, fiction, 1899 年)

また、(18) のように what が主語となる場合が 1912 年より現れている。

- (18) ...and he had promised the swamp boy to avoid doing what was bound to bring the squatters down upon them during the night.
(COHA, fiction, 1912 年)

更に、(19) のように動名詞が主語になる場合が 1940 年代より見られる。1 例のみだが、(20) のように to 不定詞が主語になる場合が 1960 年代にあった。

- (19) Moreover, regardless of precautions, borrowing under existing conditions is bound to lead to rising prices. (COHA, magazine, 1940 年)
- (20) To accept these contributions as personal gifts, however, was bound to raise black questions about vote-buying. (COHA, magazine, 1968 年)

whether 節が主語になる文や [it be bound to be Adj to 動詞] の構文は COHA では観察されなかった。⁷ 以上、初出の年代を示したが、to 不定詞の主語以外の主語は初出以降の英語でも頻繁に使われている。(13) から (20) の語彙的・統語的な特徴が時系列で発展していく様子を以下のような図で表すことができる。

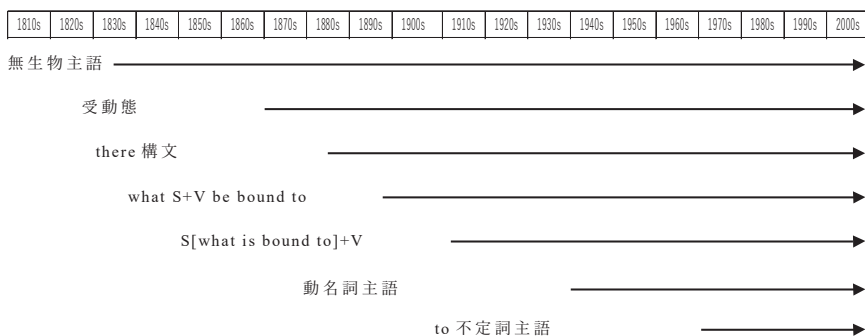


図5：be bound to の主語と統語パターンの広がり

以上、義務を表す be bound to が文法化をしながら準助動詞化する過程の中で、構文や主語の選択の可能性が広がっていく現象を COHA のデータで示した。

まず無生物主語が使われるようになることが文法化の最初の現象であると言えよう。図4で見た無生物主語の頻度が高くなる 1910 年代以前に、受動態や there 構文が使われるようになる。その後 1940 年代に頻度がピークをむかえるまでに、what 節主語、what 主語、動名詞主語という新しい統語的なパターン

の主語が出現した。1 例のみだが to 不定詞主語も 1960 年代に現れた。しかしながら、図 3 で見たように、1950 年代以降、頻度を減らしているためか whether 節や it be bound to be Adj to V 構文などの新しい統語的なパターンは見受けられず、文法化の進行としては、static な段階に入っていると言わざるを得ない。

4.2 be obliged to の推移

本項では be obliged to についても COHA に見られる傾向を明らかにする。4.1 と同じ方法を使用してデータを取った。図 6 は無生物主語の割合を示しているグラフである。表 3 がその生起数と実際の割合のデータを示している。

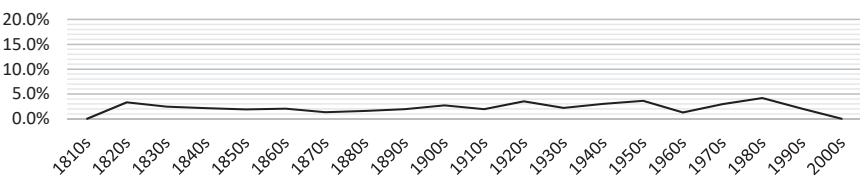


図 6 be obliged to の無生物主語の割合

表 3 be obliged to の全部の生起数、無生物主語の生起数、無生物主語の割合

	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
be obliged to 生起総数	51	419	695	655	840	777	966	747	815	697	560	591	362	265	166	156	134	119	97	72
無生物主語 生起数	0	14	17	14	16	16	13	12	16	19	11	21	8	8	6	2	4	5	2	0
無生物主語 の割合	0.0%	3.3%	2.4%	2.1%	1.9%	2.1%	1.3%	1.6%	2.0%	2.7%	2.0%	3.6%	2.2%	3.0%	3.6%	1.3%	3.0%	4.2%	2.1%	0.0%

図 6 で明らかなように、無生物主語の割合は 5% 未満である。無生物主語としてカウントした vessel や boat のように「船」を表す主語が頻繁に使用されていることが多い (COHA における無生物主語 204 例中、47 例) ののが特徴であった。実際「船」や「飛行機」がメトニミーとして使用されていると判断すれば、無生物主語の割合は更に減ることとなる。一方、抽象名詞が使用される例が be bound to に比べて少なかった。しかしながら、注目すべき統語的なふるまいが 3 点見られる。まず、受動態が 1820 年から 1904 年まで 9 例あった。(be bound to は同時期に 13 例で大きな差はない。) (21) は初出の例である。

- (21) ...it (著者注 water) was obliged to be handed from man to man in the cooking kettle, out of which the poor animals drank. (COHA, non-fiction, 1820 年)

また、(22) のような there 存在文が 1900 年と 1922 年に 2 例ある。

- (22) If you win one or two small cases, there's obliged to be undue influence of the devil. (COHA, fiction, 1900 年)

また、what を含む主語は (23) のように what が主語になる場合が 1856 年に 1 例だがある。これは be bound to の (18) の 1912 年の例より 56 年も早い。

- (23) This must stand, whatever else is obliged to give way before it. (COHA, magazine, 1856 年)

いずれも 1922 年以降は、このような例が一切なくなり、[生物主語 + be + obliged + to + 動詞] というパターンが be obliged to の主な文型となった。逆に言えば、1922 年くらいまでは、受動態で表現される義務が散見されたり、主語の選択性も制約が少ない状況であったことを考えると、be obliged to も文法化を起しつつあった可能性も考えられる。また、図 3 で見たように、1800 年代は頻度が高かったことも文法化を起しやすい環境であったと言えよう。Westney (1995: 28) は、Jespersen (1940: 323) が be obliged to は無生物主語を取ることができると言及していると記述している。Jespersen (1940: 323) は、be obliged to を一つの項目として取り上げ、例として、Austen の *Emma* (1816) に出ている this が主語になる例などを計 4 つをあげている。Jespersen は 1860 ～ 1943 年の時代を生きたが、この時代の人にとっては当時の be obliged to の用法は現在とは異なっていたと推測できる。Quirk et al. (1985: 137) が be obliged to を準助動詞にあげたのは、Jespersen (1940: 323) の記述に依拠したのかもしれない。

4.3 be bound to の今後

本項では、be bound to から派生した 2 種類の定型表現を紹介した後、今後の文法化について考えたい。1 つめは、LDOCE⁶ で定型表現としての記述がある be bound to happen である。以下が COHA での初出の例である。

- (24) One remedy for this sort of thing - which is bound to happen now and again in a system that depends so much on human agency - is to form the habit of religiously “ringing off” when you are through. (COHA, magazine, 1894 年)

COHA では総計 110 例ある。COCA には、151 例あり、be bound to の総数に対して、4.6% (151/3,307 例) がある。

2 つめは、先にも見た there be bound to である。図 7 は be bound to の総数に占める there 存在文での使用の比率を示したグラフである。表 4 は生起数と実際の割合を示している。⁸

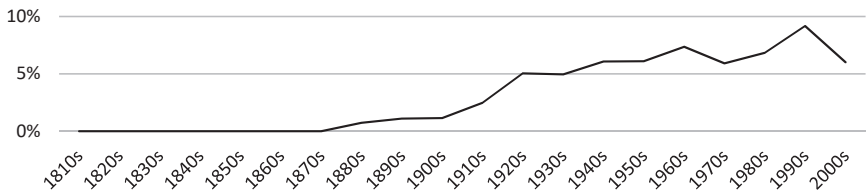


図 7 there 存在文での使用の推移

表 4 there 存在文の生起数、be bound to の生起数、there 存在文の割合

	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
there 存在文 生起数	0	0	0	0	0	0	0	2	3	4	11	22	20	28	25	28	21	19	22	12
be bound to 生起総数	22	91	220	243	232	254	304	278	277	352	448	436	404	462	411	381	356	278	240	200
割合	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	1.1%	1.1%	2.5%	5.0%	5.0%	6.1%	6.1%	7.3%	5.9%	6.8%	9.2%	6.0%

上のグラフを見ると、出現率は徐々に増えていき、1950 年以降では、7% 前後が there 存在文で使用されている。COCA のデータでは 7.4% (244/3,307 例) である。この数字は他の繰り上げ構文を起こす他のフレーズとの比率に比べると高い。例えば、繰り上げ構文の頻度が高い be likely to が使用される総数に対して同フレーズが存在文で現れる割合は、COCA において 1.2% (260/21,342 例)、be supposed to は 0.8% (228/29,682 例) である。⁹ be bound to の there 存在文での使用率の高さは際立った特徴であることは明らかである。しかしながら、本論の調査では there be bound to という定型表現をあげた辞書はなかった

ので、be bound to happen を定型表現にあげるのなら当該フレーズもあげておくべきであろう。

これら2つの定型表現だけで、COCA で be bound to の全例の内、12% がどちらかの例であることを考えると、表現する内容に多様性が乏しいと言わざるを得ないし、また、be bound to が1950年以降頻度を減らしている事実も併せて判断すると、be bound to が今後 dynamic な文法化に転向する可能性があるかという問いに関しては、否定的な見解を抱かざるを得ない。

5. 結 び

本論は be bound to の準助動詞化の過程を COHA を分析して示した。認識的用法は準助動詞化を遂げていると言えるが、義務を表す用法は、準助動詞としてはまだ完全に機能していない場合があることを示した。また、文法化の過程としては、頻度を増やしつつ、まず無生物主語が使用されるようになり、受動態、there 存在文や統語的に複雑なパターンで抽象的な主語が出現する文が現れる順序で進んだ。1950年以降は頻度を減らしつつあるのと、2つの定型表現（there 存在文での使用、be bound to happen での使用）が頻繁に使用されていることを考えると static な文法化の段階に入り、今後 dynamic な文法化に転向する可能性が少ないことを議論した。

注

- * 匿名の査読の先生方よりたくさんの貴重なコメントをいただいた。ここに謝意を表したい。いまだ残る不備な点は全て筆者が責任を負っている。本研究の成果の一部は科学研究費（基盤研究費（C）：課題番号 18K00672）の助成によるものである。
1. Quirk et al. (1985: 143) は、準助動詞 (semi-auxiliary) を 1) 助動詞 do を取らないで否定と疑問を表すが、法助動詞と異なり not は準助動詞の後に来ない（例：*is going ton't），2) 受動態や there 存在文に使用され、主語から独立している（例：Brazil is going to win the World Cup.=The World Cup is going to be won by Brazil.），3) フレーズを中心となる動詞は本動詞として to と共起して用いることができない（例：He was bound to be a failure; *Someone bound him to be a failure.），4) been の形でも、to 不定詞の中で使用可能である（例：been going to, to be bound to），5) 他の準助動詞と連鎖で使用可能である（例：Someone is going to have to complain.）と定義し、他の法助動詞や had better 等の法的なイディオム（modal idiom）と繰り上げを起こす動詞 + to（例：seem to, happen to）や本動詞 + to（例：hope to）の中間にあるものであるとしている。

2. been bound/obliged to の頻度に差があることも指摘しておきたい。COCA で、been bound to が 10 例しかないのに、been obliged to は 103 例ある、be bound to の総数は be obliged to の約 3 倍あることを考えると、be going to が have been going to のように現在完了の形になりにくのと同様に、認識的な用法では、あくまでも is, are, was, were とこれらの縮約形が基本であると言える。
3. 検索は {[vb*] * bound to [v?i*]} で行っている（レマ表示）。2 語以上の語が現れる場合の分析は割愛している。これまでの議論から義務的用法の場合は by 句がしばしば使われることから 2 語以上の場合も義務的用法が多いことが予想される。
4. be bound to は {[vb*] bound to [v?i*]}, be obliged to は、{[vb*] obliged to [v?i*]} で検索をかけている（レマ表示）。但し、個々の例の確認は行っていない。また、British National Corpus（以下 BNC）でも同様に検索をした結果、be bound to は総数 1,622 例、(spoken のサブジャンルを除くと 1,460 例)、be obliged to は総数 982 例、(spoken を除くと 959 例) というように 1.5 倍くらいである。頻度を比較すると、イギリス英語がアメリカ英語より両フレーズを頻繁に使うだけでなく、be obliged to をより頻繁に使う傾向がわかる。
5. よって表 1 と表 2 の総数は異なっている。
6. 実際の online のデータでは、its ではなく ifs という形で出ている。これはスキャンをするときに機械が誤認したものと思われる。
7. 法助動詞の will は、there 主語は 1811 年、what 主語は 1815 年、what 節主語は 1817 年、動名詞主語は 1823 年、to 不定詞は 1832 年、whether 節は 1823 年、it will be Adj to V 構文は 1817 年に出現していて、COHA のデータの早い時期にすでに散見される。
8. COHA の問題点として原文をデジタル化する際に、原文の読み取りが正確に行われていない場合がある。例えば、以下の例は he ではなく be が本来使用されていると思われる。...there is bound to he unemployment. (COHA, Fiction, 1933 年) このような場合は分析から除外している。
9. BNC のデータの場合、be bound to 8.1% (131/1,622 例) be likely to 2.3% (193/8378 例)、be supposed to 1.0% (33/3,165 例) というように、be bound to は there 存在文との共起率が高い。一方、be bound to happen は、1.6% (26/1,622 例) で、COCA に比べて頻度は低い。

コーパス

British National Corpus 2019 年 9 月～11 月アクセス

<https://www.english-corpora.org/bnc/>

Corpus of Contemporary American English 2019 年 9 月アクセス

<https://www.english-corpora.org/coca/>

Corpus of Historical American English 2019 年 8 月～11 月アクセス

<https://www.english-corpora.org/coha/>

辞書

Longman Dictionary of Contemporary English, 6th edition (2014) Harlow: Pearson Education.
Longman Dictionary of American English, 5th edition (2014) Harlow: Pearson Education.
Oxford English Dictionary, 2nd edition, version 4.0 (2009) Oxford: Oxford University Press.
『新英和大辞典』6版 (2002) 東京：研究社

参考文献

- Bybee, J. (2003) “Cognitive Processes in Grammaticalization.” In Tomasello M. (ed), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, Volume II. Mahwah: Erlbaum, pp.145–67.
- Bybee, J. (2015) *Language Change*. Oxford: Oxford University Press.
- Hopper, P. and E. C. Traugott. (2003) *Grammaticalization*, Second edition. New York: Cambridge University Press.
- Jespersen, O. (1932) *A Modern English Grammar: On Historical Principles*, Part IV. London: George Allen & Unwin.
- Jespersen, O. (1940) *A Modern English Grammar: On Historical Principles*, Part V. Copenhagen: Ejnar Munksgaard.
- Mair, C. (2004) “Corpus Linguistics and Grammaticalization Theory: Statistics, Frequencies, and Beyond.” In Lindquist, H. & C. Mair. (eds), *Corpus Approaches to Grammaticalization in English*. Amsterdam/Philadelphia: John Benjamins, pp. 121–150.
- Mong, K. Y. (n.d.) “A Semantic Analysis of *be prepared/ determined/ fated/ bound/ required to*: Desemanticization and Emancipation.” *Academia*. file:///C:/Users/Owner/AppData/Local/Microsoft/Windows/INetCache/IE/K449DXWG/A_semantic_analysis_of_be_prepared_deter.pdf 2019年10月アクセス
- Palmer, F. R. (1990) *Modality and English Modals*, Second edition. London & New York: Longman.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. (1985) *A Comprehensive Grammar of English Language*. London & New York: Longman.
- Westney, P. (1995) *Modals and Periphrastics in English*. Tübingen: Niemeyer.

「研究ノート」

The Expansion of the *how about* Construction in American English

Satoshi YAMAZAKI

Abstract

This paper examines the recent spread of the construction in which the idiomatic phrase *how about* is followed by a sentence (*how about S*) and of those patterns followed by other categories considered to be syntactically incompatible with the preposition *about* in American English. We argue that the gradual advancement of the chunking of *how about* in the sense proposed by Bybee (2007; 2010), together with the entrenchment of its frequent instances/exemplars as independent chunks, is responsible for the complement expansion of *how about*.

1. Introduction

The *how about* (an accusative or possessive pronoun (*Acc/Poss*)) *Ving* construction, as in (1a) and (1b), quoted from the Corpus of Contemporary American English (COCA), has been employed to suggest or direct a certain course of action to the addressees¹; however, in the last decade or two, the construction in which *how about* takes a sentential complement (hereafter, *how about S*) instead of *Ving* has been gaining ground, especially in informal American English, as in (1c):

- (1) a. Well, **how about having** supper with me tonight? [*How about Ving*]
(COCA, Fiction, 2010s)
- b. “Will you shut up?” “**How about you shutting** up?” [*How about Acc Ving*]
(COCA, Fiction, 1990s)
- c. “**How about you invite** her over to dinner this weekend?” [*How about S*]
(COCA, Fiction, 2010s)

A notable aspect of *how about S* is that under the traditional grammatical stance, the combination of the preposition *about* followed by a sentence is, in general, syntactically anomalous (Quirk, et al., 1985: 658f.). Furthermore, examination of corpora indicates that other categories considered to be syntactically incompatible with *about* have also increasingly been combined with *how about* in recent decades.

This paper examines the recent expansion of these irregular complements of *how about* and argues that “chunking,” as defined by Bybee (2007; 2010), has played an important role in the growing popularity of *how about S* as well as other types of complements incompatible with *about*.

Chunking, according to Bybee (2010), is the linguistic process where adjacent morphemes or words that are repeatedly used together come, in time, to be stored and accessed as single units or chunks. Under this view, compounds, idioms, constructions and even constituent structures are units “chunked” through repetition: chunking is the driving force in their creation. Because chunks are stored/accessed as wholes, they become autonomous, and their compositionality and/or internal structures may be more or less lost.

Let us briefly examine this chunking process with the formation of the complex preposition (CP) *in spite of*, as explained by Bybee (2010: 138–146). While the noun *spite* in this CP was employed with the meaning of “contempt or defiance,” the CP remained compositional because the relation between this noun and other noun instances was transparent. However, as the CP later developed the counter forces and concessive meanings, as found in *in spite of all opposition*, the noun grew remote from other instances. The repetitive use with these meanings led to the decategorization of the noun; the entire phrase became more autonomous, and is now largely accessed as a single unit (see also Hoffmann, 2005, chap. 3) or as a chunk. Chunking also plays an important role in Maeda’s (2019) analysis in which he argues that “half-gerunds” such as *be busy Ving* and *have trouble Ving* have developed from their gerundial counterparts with the preposition *in* through constructionalization.

The paper is structured as follows: Section 2 examines the spread of *how about S* in two corpora from the BYU suite. Section 3 illustrates that in the *how about* construction, there are other types of anomalous syntactic combinations that have been increasing in recent decades. Section 4 outlines Bybee’s (2007; 2010) notion of chunking and argues that advanced chunking, together with the entrenchment of

frequent instances as independent chunks, is responsible for the expansion of the complement types for *how about*. Section 5 offers a brief summary.

2. The spread of *how about S*

This section examines the spread of *how about S* in different text genres, both in recent decades and in a somewhat longer diachronic span in some detail because this aspect is the most conspicuous in the recent complement expansion of *how about*. To examine this spread, we address the rivalry between the *how about S* and *how about (Acc/Poss/NP) Ving* constructions as they both contain predication and in this sense are comparable.² We do not consider the *how about if* construction in relation to this rivalry because this construction is sometimes used to ask what should be done if something happens.

Let us first look at the recent spread of *how about S* in different text genres in COCA (last accessed September 2019). Figures 1a to 1c show the rivalry between *how about (Acc/Poss/NP) Ving* and *how about S* in newspapers (Figure 1a), spoken English (Figure 1b), and works of fiction (Figure 1c) in three five-year periods, namely, 1990–94, 2000–04, and 2010–14, respectively. Each figure shows the ratio of *how about (Acc/Poss/NP) Ving* to *how about S* in every ten-year span. The normalized frequencies (per million words (pmw)) are provided on each bar. The search strings employed for *how about (Acc/Poss/NP) Ving* are “how about (*) _v?g*” ((*) indicates with or without one intervening word hereafter), which retrieved such patterns as *how about + pronoun/just Ving* as well, and “how about _pp* * _v?g*” (_pp* represents any personal pronoun), which returned one relevant attestation. The strings of “how about _p* (*) _vv0*” (_p* represents any pronoun and _vv0* base forms of any verb) were employed for the retrieval of the *how about S* attestations.³ As may be expected, there are a few examples that are not captured by these queries. However, a manual checking of the same queries for all the complement types of *how about* in the Corpus of Historical American English (COHA) (see Section 3) revealed that the vast majority of the relevant attestations were, in fact, covered by these queries. Such irrelevant attestations as *How about bullfighting*, for instance, were manually excluded from the results.

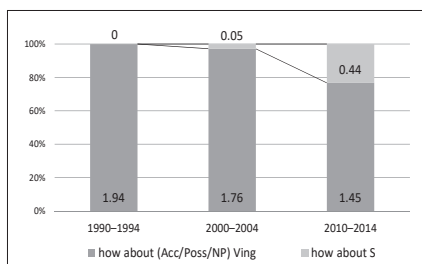


Figure 1a: The rivalry between *how about (Acc/Poss/NP) Ving* and *how about S* [COCA, Newspapers]

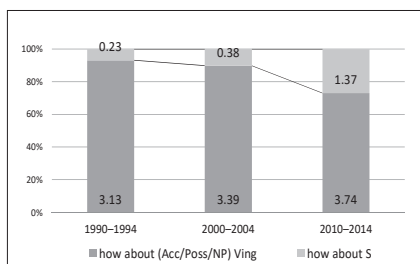


Figure 1b: The rivalry between *how about (Acc/Poss/NP) Ving* and *how about S* [COCA, Spoken]

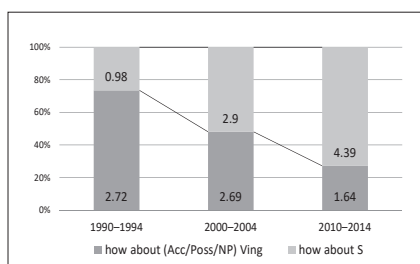


Figure 1c: The rivalry between *how about (Acc/Poss/NP) Ving* and *how about S* [COCA, Fiction]

While it is apparent that the proportion of *how about S* as compared to *how about (Acc/Poss/NP) Ving* has been more or less increasing across text genres, its spread in the Fiction genre is remarkable: in just two decades, the ratio of the two constructions has almost been reversed. While the informality of conversation is a matter of degree, virtually all attestations in the Fiction genre are deemed to belong to fictional casual conversation, which seems to be the more commonly favored register of *how about S*. At first glance, its relative scarcity in the Spoken genre may be surprising. However, this is probably due to the fact that the Spoken section of COCA comprises public broadcasts such as news broadcasts/analyses, talk shows, and interviews with politicians and celebrities. It is perhaps to be expected that these talks tend to be more formal than the casual conversations which many conversations in works of fiction try to depict.

What has been the trajectory of informal spoken English? Although we could not find a corpus of authentic daily conversation in American English large enough to serve

our purposes, the Movie Corpus provided a wealth of examples. This corpus also contains data from the early 20th century. Figure 2 is the result of the analysis of the attestations in American movies in the Movie Corpus (last accessed March 2020). Attestations in Canadian movies and those jointly produced with the USA are manually excluded. The normalized frequencies (pmw) of occurrences are shown. Because a larger number of attestations of *how about Acc/Poss Ving* were found, the data are provided separately. The queries were performed with the basic patterns of “how about _v?g*” for *how about Ving*, “how about _p* _v?g*” and “how about _app* _v?g*” (_app* represents possessive personal pronouns) for *how about Acc/Poss Ving*, and “how about _p* _vv0*” for *how about S*.

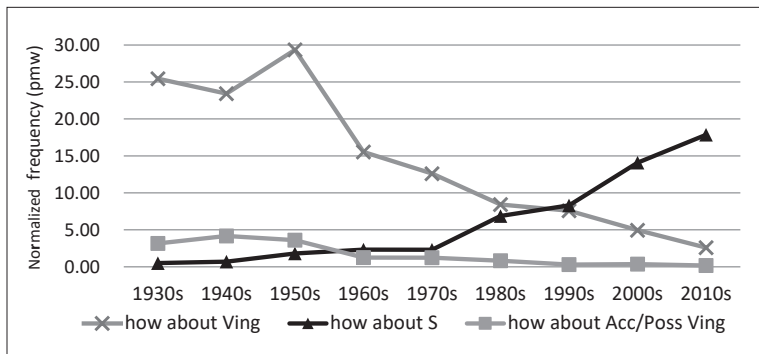


Figure 2: The rivalry among 3 constructions
[The Movie Corpus, American movies only]

An examination of the diachronic rivalry of the three constructions reveals some interesting points. The frequency of *how about S* started increasing in the 1980s and soon surpassed that of the declining *how about Ving*. In the Movie Corpus, the use of *how about S* (17.87) is nearly seven times more frequent than the use of *how about Ving* (2.58), which represents a much wider gap than we saw in the Fiction section of COCA (see Figure 1c). Therefore, it seems reasonable to consider that in informal conversation, *how about S* is now far more likely to be employed than *how about Ving*. It can also be seen that until the 1950s, *how about Acc/Poss Ving* was a more common choice than *how about S*, although it is now rare in informal English.

3. Other syntactically unusual complements of *how about*

The previous section discussed the recent spread of *how about S* in different text genres in American English. One unusual aspect of this construction, however, is that it has the anomalous syntactic combination of the preposition *about* followed by a sentence (see Section 1). Taylor (2012) offers *what about* as an example of “syntactic or constructional idioms,” which have unusual internal syntactic structures. He points out that *what about* is followed by such various items as “noun phrases, prepositional phrases, *V*ing phrases, and certain kinds of subordinated clauses (. . . *if*, *when*, *where*, *before*, and *after*, for example. . .)” (p. 87). *How about* is also followed by these items. Although Taylor does not mention this, it would also be syntactically unusual for some of these items, that is, PPs and adverbial subordinate clauses, to follow the preposition *about*. Examples of *how about* followed by categories that are considered to be syntactically unusual are provided from COHA in (2):

- (2) a. **How about if** I come back and get you at about one-thirty?
[*how about* + *if*-clause]⁴ (COHA, Fiction, 1990s)
- b. “That’s putty well, when a man hes any worldly goods” said Sam; “but **how about when** he hesn’t?” [*how about* + *when*-clause] (COHA, Fiction, 1860s)
- c. ISABEL: There’s only one guy for her. IAN: **How about for** you?
[*how about* + PP] (COHA, Fiction, 2000s)
- d. [After watching a movie] Linda said, “Now, wasn’t that [i.e., movie] cute?”
“**How about silly**,” I said. “That’s almost like cute.” [*how about* + Adjective]
(COHA, Fiction, 1980s)
- e. “Can we make it earlier? I have to work then.” . . . “No.” . . . “**How about later**?” “No.” [*how about* + Adverb] (COHA, Fiction, 2000s)
- f. “You don’t say something?” “Like what?” I asked reluctantly. She shrugged, still smiling. “**How about thank you**?” [*how about* + adverbial formulaic expression] (COHA, Fiction, 2000s)

In this section, we first briefly examine COCA to confirm that the combinations of *about* + PP are generally unusual. We do not discuss the irregularities of the other combinations exemplified in (2) because they seem to be obvious and do not require

further examination.

It is generally believed that only certain prepositions take PP complements. Huddleston and Pullum (2002: 640), for instance, point out that *from*, *since*, *till/until* “most readily take PP complements,” as exemplified by *He emerged **from under** the bed* or ***until after** six*. The preposition *to*, on the other hand, does not generally take PP complements except in some combinations. Brown and Miller (2016: 176) likewise state that most prepositions do not take PP complements.

Now, we examine the extent to which the abovementioned [P + PP complement] patterns are represented in the Fiction and Spoken sections of COCA (last accessed August 2019). There were more than 3000 occurrences of *from under*, more than 5000 of *from behind*, 420 of *until after*, and 101 of *since before*, excluding those in which *after* in *until after* and *before* in *since before* were used as subordinating conjunctions. Then, with regard to the pattern of [*about* + PP complement], searches were conducted with the string of [_n* about _i*] (any noun + *about* + any preposition).⁵ No relevant examples were found in the Fiction section, but two possible examples were found in the Spoken section, both of which are shown in (3):

- (3) a. BARBARA–ALTERS (Off-camera) You are very different off stage than on. What’s the difference? JOAN RIVERS (COMEDIENNE) On stage, I say all the things **about in the shower**. (COCA, Spoken, 2010s)
- b. Michael, you criticized me for putting out just one small idea **about after school**. (COCA, Spoken, 1990s)

It can be seen, however, that in (3a), something is omitted in the [P + PP] phrase, such as in *about what I say (or sing) in the shower*. As for (3b), *after school* is not a usual compositional PP but is rather accessed as a single unit, as suggested by the existence of its derived adjectival compound, *after-school* (club). This may make it possible for *about* to take this particular PP. Thus, it seems that there is something special even with these rare attestations.

As we have confirmed that it is syntactically unusual for *about* to be followed by PPs, Table 1 is provided to show all the attestations of *how about* in COHA, grouped by their complement types (last accessed December 2019).⁶ The category “Others” includes *how about* followed by adjectives, adverbs including *here* and *there*,⁷ and other

miscellaneous categories, as in (2f).

Table 1: The distribution of the *how about* attestations according to their complement types [COHA, all text genres]

	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	Totals
how about NP	8	9	35	49	57	90	134	320	380	396	390	488	505	465	410	504	425	4665
how about (Acc/Poss/NP) Ving	0	0	2	1	3	3	13	27	25	74	79	101	80	66	54	62	58	648
how about wh-interrogative	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	4
how about when	0	0	1	0	0	0	0	1	2	1	0	1	8	1	0	1	4	20
how about if	0	0	0	0	0	0	1	0	0	0	2	0	8	5	12	15	21	64
how about S	0	0	0	0	0	0	0	0	1	0	0	5	5	4	16	25	55	111
how about PP	0	0	0	0	0	0	0	0	0	0	1	3	7	3	7	5	7	33
Others	0	0	0	0	0	0	0	0	0	0	2	4	5	5	9	14	15	54
Totals	8	9	38	50	60	93	148	348	408	471	474	603	618	549	509	627	586	5599
Percentages of unusual complements	0.00%	0.00%	2.63%	0.00%	0.00%	0.00%	0.68%	0.29%	0.74%	0.21%	1.05%	2.16%	5.34%	3.28%	8.64%	9.57%	17.06%	

*The total number of occurrences of *how about Acc/Poss/NP Ving* is 85.

The complement types observed can be divided into two groups: those compatible with the preposition *about* as complements, i.e., NPs, Vings and *wh*-interrogatives, and those incompatible or unusual, i.e., adverbial *when/if*-clauses, sentences, PPs and others (shaded instances). The question arises as to how *how about* came to take these apparently contradictory complement types. In the next section, we argue that this was made possible because of the advancement of chunking of *how about*, as proposed by Bybee (2007; 2010), on one hand, and the entrenchment of frequent instances, on the other.

4. Gradual chunking of *how about* and the entrenchment of frequent instances as chunks

As already mentioned in Section 1, chunking is the linguistic process where adjacent morphemes or words that are repeatedly used together come, in time, to be stored/accessed as single units or chunks. “Items that are used together fuse together.” (Bybee, 2007: 316; see also Krug, 1998). As chunks are repeatedly accessed as wholes, they become more autonomous, which is more obvious in the case of chunks formed by phonetic reduction such as *I’m* and *you’ll*; those which underwent semantic/pragmatic shifts such as *in spite of* (see Section 1), the discourse marker *indeed*, or the idiom *spill the beans*; or those in which both of the above processes were involved,

such as in derived words like *disease* and *business*. The compositionality and/or analyzability/internal structure of those chunks are more or less lost (Bybee, 2010, chap. 3).

Bybee (2010) also points out the gradient nature of chunks as follows:

[T]he status of a chunk in memory falls along a continuum. . . . [T]here is a continuum from words that have been experienced together only once and fairly recently, which will constitute a weak chunk whose internal parts are stronger than the whole, to more frequent chunks such as *lend a hand* or *pick and choose* which are easily accessible as wholes while still maintaining connections to their parts. . . . On the high-frequency end of the continuum, chunks such as grammaticalizing phrases or discourse makers do lose their internal structure and the identifiability of their constituent parts[.] (p. 36)

This characterization concerns the synchronic gradient nature of chunks, but it is readily translated into diachronic chunking processes, as can be seen in Bybee's (2010) analysis of *in spite of*, outlined in Section 1. With regard to *how about*, we argue that it has gone through two stages of chunking. Specifically, while almost all instances of *how about* in COHA were either *how about NPs* or *how about Vings* until the middle of the 20th century (see Table 1), *how about* was a "weaker chunk" whose internal parts were still identifiable, corresponding to Bybee's second type of chunks mentioned above. Bybee (2010: 36) points out that since the internal structure of "more frequent chunks" (weaker chunks in our terms) such as *lend a hand* is identifiable, a modifier or an indirect object can be added, as in *lend a helping hand* or *lend me a hand*. It is true that *how about* has been a fixed idiom disallowing modification, but given the fact that virtually all instances were *how about NPs/Vings*, its component parts including the preposition *about* were considered to be still identifiable at this stage.

Then, as *how about* continued to be used repeatedly, its internal structure was gradually blurred and then lost. *How about* was now stored/accessed as a (perhaps adverbial) holistic chunk ("fixed chunk" at this stage hereafter). As a result, *how about* was ready to take S's and other "complements" apparently incompatible with the preposition *about*.⁸

As for the approximate period when *how about* developed from a weaker into a fixed chunk as shown in Table 1, the number of shaded attestations in Table 1 is very

low until the middle of the 20th century (see also “Percentages of unusual complements” at the bottom of Table 1). While the transition into a fixed chunk was apparently under way during this period, a substantial change occurred in the latter half of the 20th century. In narrowing this down, the change in the number of occurrences of *how about Ving* is important because this form is sensitive to the internal structure of *how about* including the preposition *about*. *How about NP*, in this respect, can be regarded as licensed by either a weaker or a fixed chunk: NPs can be syntactically followed by *about*, or they are expected to be semantically compatible with *how about* as a holistic chunk (see the next paragraph for its treatment). In Figure 3, these elements have been considered, and it displays the rivalry between *how about (Acc/Poss/NP) Ving* and the shaded constructions of *how about if*, *how about S*, and combined *how about when/PP/others* in Table 1. The graph is plotted based on the normalized frequency of pmw.

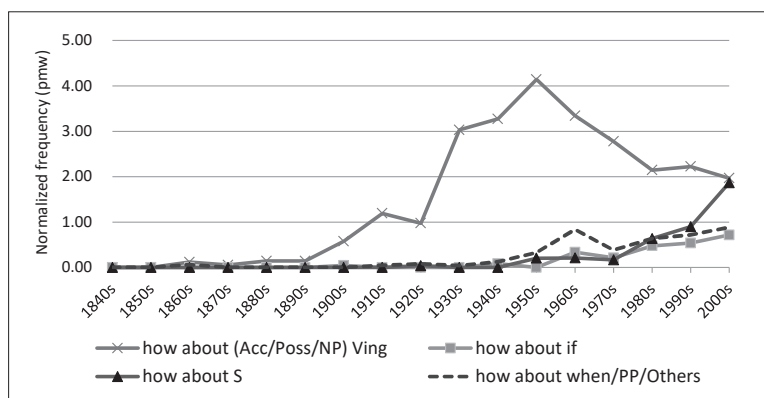


Figure 3: The rivalry between *how about (Acc/Poss/NP) Ving* and shaded constructions in Table 1 [COHA, all text genres]

Taking into account the overall increase in the attestations of three sets of unusual complements, on one hand, and the decrease in those of *how about (Acc/Poss/NP) Ving*, on the other, we assume that it was probably in the 1980s or 1990s that *how about* came to be stored/accessed primarily as a fixed chunk.

This assumption is apparently problematic because even after the 1980s–90s, *how about Ving*, which is supposed to be sensitive to the internal structure of *how*

about, remains a dominant pattern. However, it has been pointed out that frequent items such as irregular verb forms are entrenched as they are, resisting analogical leveling (Bybee, 2007: 271). *How about Ving* became so frequent after the 1930s that it is likely that this pattern itself had come to be stored/accessed as a chunk with the open slot of *Ving* by the 1980s. Because of its very high frequency, *how about NP* is also considered to have been stored as an independent chunk with the open slot of NP by the 1980s.^{9,10} Thus, the mode of access to *how about Ving* and *how about NP* changed by this time gradually and, maybe, with a certain overlapping period. In contrast, *how about Acc/Poss/NP Ving* has never had a high enough frequency to be stored as an independent chunk (see note in Table 1). It is supposed to be licensed by *how about* as a weaker chunk, for which *how about* as a weaker chunk is still occasionally activated.

Figure 4 summarizes the assumption of the gradual chunking that *how about* has gone through. The darker shades on the bars indicate the major mode of access to *how about* and the lighter ones (extremely) the minor mode of access on the timeline. The two oval items indicate the independently stored chunks with their tentative starting periods shown.

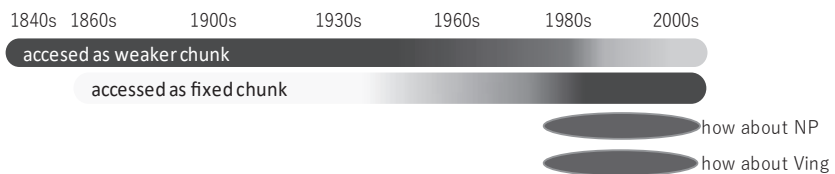


Figure 4: Two stages of the chunking of *how about* and the entrenched independent chunks

Based on Figure 4, *how about* may be accessed as either a weaker or a fixed chunk during the same period, although it is strongly biased to either mode of access. We consider that this synchronic instability in the access mode is natural rather than unusual, considering that language change is, in general, a gradual process.¹¹

Finally, we address the issue of *how about* as a fixed chunk because of the loss of its internal structure, erroneously licensing constructions that are in fact (very) rare or virtually impossible as its complements. One such construction type is bare or *to* infinitives.¹² The reason they are (very) rare may be attributed to the existence and saliency of the entrenched independent chunk of *how about Ving*: when people

combine a VP element with *how about*, this chunk is activated—thus, infinitive patterns are basically blocked. Another complement is *for-to* infinitives. It is noticed, however, that this construction is functionally parallel with *how about S* in that they both comprise the subject and its predicate; therefore, they come into rivalry. Evidently, the subject + base verb form of *how about S* is much simpler and easier to employ than *for-to* infinitives, which is probably the reason the latter construction is avoided.

5. Conclusion

This paper has examined the spread of *how about S* and other categories that are apparently at odds with the complements of *about* in American English, and we have argued that the advanced chunking and the entrenchment of frequent instances as independent chunks are responsible for the expansion of its complement types in recent decades.¹³

Acknowledgements

I would like to thank three anonymous reviewers for their helpful suggestions and detailed comments on earlier versions of this paper. Needless to say, all remaining inadequacies are my own.

Notes

1. Needless to say, *how about Ving* is also occasionally used to ask about another aspect of what is being talked about (The *Macmillan English Dictionary for Advanced Learners*, s.v. *how*); *how about S*, however, is very rarely employed in this sense (Denawa, 2017).
2. Note that *how about NP Ving* is rare. Moreover, this paper considers the (*Acc/Poss/NP Ving* complement of *how about* (at least as a “weaker chunk” (see Section 4)) a gerund, which is indicated by the fact that the subject may be in the possessive case (Huddleston and Pullum, 2002: 1220).
3. Note that the subject of *how about S* is generally limited to personal pronouns.
4. Although *how about if* may be now established as an autonomous idiom, it is considered that originally the adverbial conditional *if*-clauses were compositionally combined with *how about* to form this idiom.
5. These searches were performed because those with [_v* about _i*] returned too many occurrences to check.

6. Identical attestations that occur more than once are counted as one. A handful of attestations of false starts, which are abandoned halfway in fictional conversations, are excluded from the table.
7. Such temporal words/phrases as *tomorrow*, *today*, *now*, *(this/next) Sunday*, and *this afternoon* are not counted as adverbs but as NPs in Table 1. This is because they can stand as the subjects of copula sentences, as in *Tomorrow/Today/Next Sunday is my birthday*, and they may be employed as NPs in the *how about* construction.
8. Regarding the relation between the chunking of *how about* and its status as an idiom, we assume that *how about* was already established as an idiom or a weaker chunk when it was first attested in COHA because it had already developed its conventionalized meanings with its fixed, syntactically irregular form. However, this idiomatic phrase has, in our view, further lost its internal structure and developed into a fixed chunk; in other words, its idiomaticity has increased.
9. It is considered that chunks may contain open slots of Ving as exemplified by such idioms or constructions as *look forward to Ving* and *there is no Ving*; these Vings are presumably parts of chunks stored as wholes. As for the open NP slot, it is contained, for instance, in the *what's X doing Y* construction (Kay and Fillmore, 1999) and the resultative *drive X crazy* construction (Bybee, 2010).
10. It is quite possible that *how about NP*, due to its high frequency, has been stored as a chunk since an earlier period. However, as there is no available evidence that suggests a particular period, this paper assumes that it has been entrenched as an independent chunk by the 1980s.
11. An anonymous reviewer highlighted the gap between the chunking of *in spite of* and that of *how about* suggested in this paper: while the former chunking largely resulted in the change in the kind of the NP complements of the CP and its semantic change, the latter has led to a categorical shift when there was no semantic change in *how about*. While this is an important point to consider, we just want to emphasize here that they differ in that the chunking of *in spite of* results in the same category as *of* but that of *how about* inevitably ends up in a different, perhaps adverbial, element.
12. I thank the anonymous reviewer who reminded me of this point.
13. This line of argument is considered to fare well with the complement expansion of idiomatic *far from* as revealed in Akimoto's (2000) study.

References

- Akimoto, M. (2000) "How far has *far from* become grammaticalized?" In Brinton, L. J. (ed.), *Historical Linguistics 1999*. Amsterdam/Philadelphia: John Benjamins, pp. 1–11.
- Brown, K. and J. Miller (2016) *A Critical Account of English Syntax: Grammar, Meaning, Text*. Edinburgh: Edinburgh University Press.
- Bybee, J. (2007) *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

- Bybee, J. (2010) *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Denawa, T. (2017) “Setsu o shitagaeru how about nitsuite [On *how about* taking clauses],” *The Journal of Toyo University Graduate School* 54: 279–290.
- Davies, M. (2008–) *The Corpus of Contemporary American English (COCA): 600 million words, 1990–present*. Available online at <https://www.english-corpora.org/coca/>.
- Davies, M. (2010–) *The Corpus of Historical American English (COHA): 400 million words, 1810–2009*. Available online at <https://www.english-corpora.org/coha/>.
- Davies, M. (2019–) *The Movie Corpus: 200 million words, 1930–2018*. Available online at <https://www.english-corpora.org/movies/>.
- Hoffmann, S. (2005) *Grammaticalization and English Complex Prepositions: A Corpus-based Study*. Oxon and New York: Routledge.
- Huddleston, R. and G. K. Pullum (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Kay, P. and C. J. Fillmore (1999) “Grammatical constructions and linguistic generalizations: the *what’s X doing Y?* construction,” *Language* 75, 1: 1–33.
- Krug, M. (1998) “String frequency,” *Journal of English Linguistics* 26, 4: 286–320.
- Maeda, M. (2019) “Han-domeishi no hattatsu to kobunka [The development of half-gerunds and constructionalization],” *Studies in Modern English* 35: 59–84.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Taylor, J. R. (2012) *Mental Corpus: How Language is Represented in the Mind*. Cambridge: Cambridge University Press.

(山崎 聡 千葉商科大学 Email: s2yamaza@cuc.ac.jp)

英語コーパス学会 第45回大会

【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）

	司 会	石井 康毅（成城大学）
1. 会長挨拶		投野由紀夫（東京外国語大学）
2. 開催校挨拶		五百藏高浩（高知県立大学副学長）
3. 総会		
4. 学会賞審査報告		西村 秀夫（三重大学）
5. 事務局からの連絡		

〈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

—クラスター分析—

後藤 克己（中部大学大学院生）

研究発表5 15:50-16:20

Helsinki Corpus ME3-EModE3 での binomials の史の変遷

—特に頻度と語源から

谷 明信（兵庫教育大学）

〈休 憩 16:20-16:30〉

〈講 演 16:30-18:00（場所:A101）〉

司 会 投野由紀夫（東京外国語大学）

Revisiting A Corpus—the BNC 2014

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

—(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）

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）〉

人工知能とデーター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 Schimd (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), 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.

■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「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。

【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、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), 「個性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 (Ø) 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一般的なである。

(1) PDE における使役動詞 *make* と不定詞

- a. They **made him sign** the contract 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))

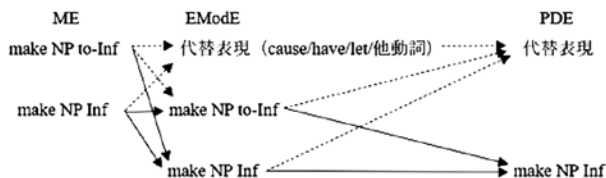
しかし、通時的に見ると態による不定詞の選択制限はなかったとされる。

(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*] *made 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.

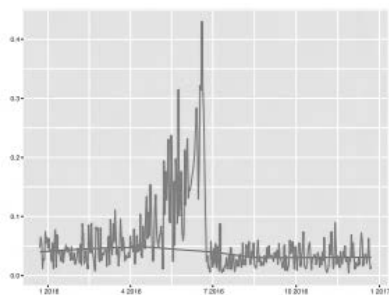


Figure 1: A topic identified via LDA for the word *Brexit* (label: “Campaign for the Brexit referendum”)

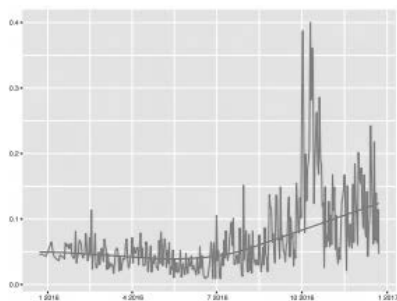


Figure 2: A topic identified via LDA for the word *referendum* (label: “Economic uncertainty associated with the referendum results”)

【研究発表第4セッション】

【研究発表4】

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）

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度となる“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。

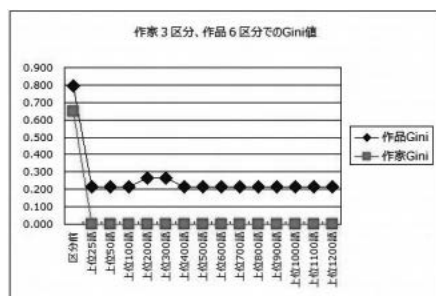


図2 2.5乗根圧縮データによる区分性

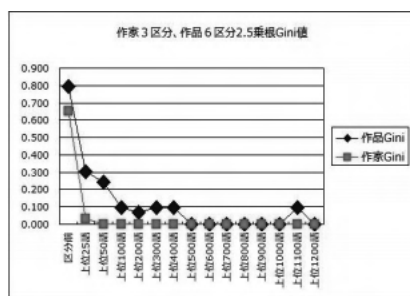


図1 圧縮なしデータによる区分性

【研究発表5】

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)

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 Kopaczkyk & 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の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りと考えられる。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。

表 1：n-gram パターン

3	4	5	6
*_A key_J r ole_N	*_A key_J r ole_N *_I ⁽¹⁾	*_V *_A key_J r ole_N *_I	*_V *_A key_J r ole_N *_I *_A
			*_V *_A key_J r ole_N *_I *_V
		*_A key_J r ole_N *_I *_V ⁽²⁾	*_V *_V *_A key_J r ole_N *_I
		*_A key_J r ole_N *_I *_A	*_V *_A key_J r ole_N *_I *_V ⁽³⁾
	*_V *_A key_J r ole_N	*_V *_A key_J r ole_N *_I *_A	*_V *_A key_J r ole_N *_I *_A
		*_V *_A key_J r ole_N *_I	*_V *_A key_J r ole_N *_I *_A
			*_V *_A key_J r ole_N *_I *_V
		*_V *_V *_A key_J r ole_N	*_V *_V *_A key_J r ole_N *_I

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

番号	具 体 例
(1)	a_A key_J r ole_N in_I (35548), the_A key_J r ole_N of_I (1135), a_A key_J r ole_N for_I (675)...
(2)	a_A key_J r ole_N in_I help_V (1415), a_A key_J r ole_N in_I develop_V (511), a_A key_J r ole_N in_I ensure_V (459)...
(3)	play_V a_A key_J r ole_N in_I help_V (1335), play_V a_A key_J r ole_N in_I develop_V (475)...

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、自由作文を課することが多い。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うによって、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。

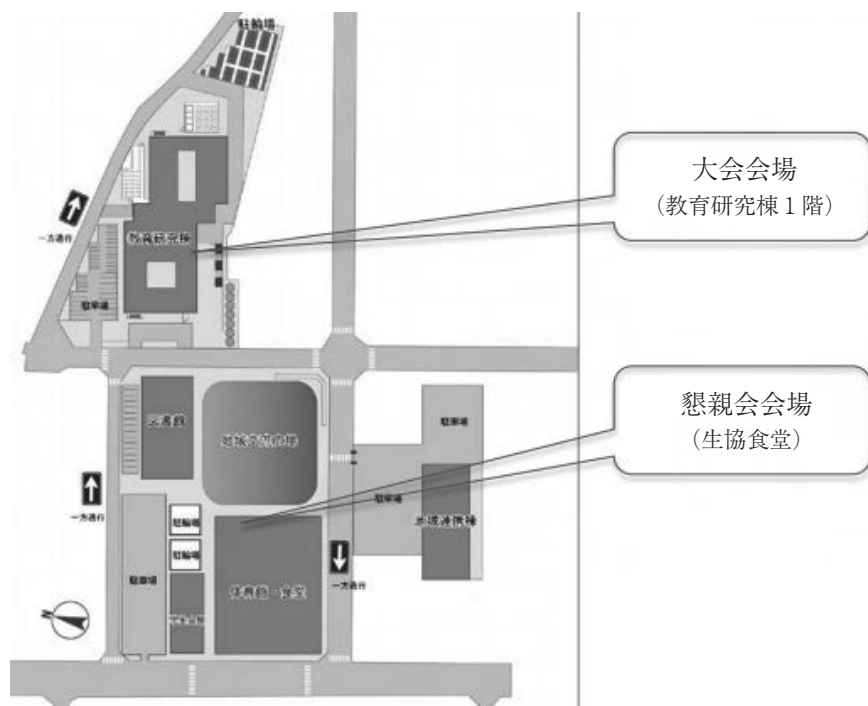
【北はりまや橋から】

- ・ 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、お問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。

『英語コーパス研究』 投稿規定

(2019年10月改定)

1. 投稿資格

投稿は会員に限る。共著の場合、第一著者は会員であることとし、その他の共著者については会員でなくてもよい。

2. 原稿の種類と長さ

【研究論文】

英文 A4サイズ 1ページあたり35行（文字数の指定はしない）、周囲の余白1インチ（25.4mm）、17ページ以内（Times New Roman 10.5ポイント使用）

和文 A4サイズ 1ページあたり35行（文字数の指定はしない）、投稿時17枚以内（明朝体フォント（游明朝・ヒラギノ明朝など）10.5ポイント使用）

※和文中の英文のフォントについては Times New Roman を原則とする。Century は用いてはならない。

（いずれも Abstract（英文300語以内）、図表、注、参考文献目録、付録、謝辞、著者情報などを含む。）

【研究ノート、総説論文・書評論文（Review article, Book review）】

・研究ノート：論文のカテゴリーに属さない小論文や萌芽的な研究、新しい研究開発の成果などをまとめたもの

・総説論文：体系的かつ網羅的に先行研究をまとめたもの

・書評論文：専門書の研究分野への貢献と課題点を明確にしたもの

英文 A4サイズ 1ページあたり35行（文字数の指定はしない）、周囲の余白1インチ（25.4mm）、12ページ以内（Times New Roman 10.5ポイント使用）

和文 A4サイズ 1ページあたり35行、投稿時12枚以内（明朝体フォント（游明朝・ヒラギノ明朝など）10.5ポイント使用）

※和文中の英文のフォントについては Times New Roman を原則とする。Century は用いてはならない。

（いずれも Abstract（英文300語以内）、図表、注、参考文献目録、付録、謝辞、著者情報などを含む。）

【その他（ソフトウェアレビュー、書評（図書紹介）、コーパス紹介など）】

研究論文の半分以内の分量

3. 原稿作成時の注意

下記のように投稿者を特定できるような情報、その他、本人の同定につながると考えられる情報は、採用決定後の最終原稿に追記するものとし、投稿時には記載しないこと。

- (1) 謝辞など
- (2) 「本論は、英語コーパス学会第X回大会において口頭発表した内容に加筆修正を施したものである。」などの文言
- (3) 「筆者が収集し、WWW (<http://...>) で公開しているデータ…」など、筆者情報につながる URL 情報など
- (4) 「拙論 (2006) で論じたように…」などと記して、参考文献目録で当該文献を参照している場合、「拙論」ではなく著者 (2006) として表記すること。

4. 提出方法など

- (1) 下記の (A) 原稿ファイル (Microsoft Word で作成したファイルとその PDF ファイル)、(B) 著者情報ファイル、(C) 論文投稿チェックシートの3種類のファイルを電子メール添付で提出。(B)、(C) については Web 掲載のフォーマットを使用のこと。
- (2) 電子メールの件名 (Subject) は「『英語コーパス研究』投稿原稿 (著者氏名)」とすること。
- (3) 提出先、締め切り期日等に関しては学会 Web サイトを参照のこと。

(A) 原稿ファイル

- a. 提出するファイル名は「原稿題名 (著者氏名)」とすること。
- b. 原稿題名の前に「論文」、「研究ノート」、「総説論文」、「書評論文」、「コーパス紹介」などの種類を明記すること。
- c. 原稿本体の冒頭には上記種類の別と題名のみを記すこと。

(B) 著者情報ファイル：「著者情報 (著者氏名)」

- a. 和文原稿の場合は英文タイトル、英文原稿には和文タイトル
- b. 著者氏名 (ふりがな・ローマ字表記)
- c. 所属
- d. 郵便番号・住所・電話番号
- e. 電子メールアドレス

(C) 論文投稿チェックシート：「論文投稿チェックシート (著者氏名)」

Web 掲載のチェックシートの必要項目すべてに ☒ を入れること。

5. スタイル

投稿論文は，研究論文，研究ノート，総説論文・書評論文の別，また，和文・英文の別にかかわらず、『英語コーパス研究』スタイルシートに従い執筆することとする。

6. 掲載論文等の電子化

掲載された論文等の著者は，論文等を電子化して学会ホームページで公開することに同意する。

7. 著作権

掲載された論文等の著作権は，本学会に帰属する。本学会は掲載論文等を印刷媒体・電子媒体で公開する権利を有するものとする。ただし，著作者が自著論文等を自分のホームページに掲載したり，自著の本に転載したりすることは妨げない。

8. 研究倫理

投稿にあたっては，下記文書などを参照し，不正行為のないようにすること。
独立行政法人科学技術振興機構『研究者のみなさまへ～研究活動における不正行為の防止について～』<https://www.jst.go.jp/contract/kisoken/h25/others/h25s805others131120.pdf>

英語コーパス学会会則

第1条 本会は「英語コーパス学会」(Japan Association for English Corpus Studies, 略称 JAECS) と称する。

第2条 本会は英語コーパス及びコーパスツールの開発・評価・利用に関わる研究, また, 英語コーパスを用いた言語研究・言語教育研究・関連研究を促進することを目的とする。

第3条 本会は前条の目的を達成するために, 次の事業を行う。

- (1) 大会・研究会等の開催
- (2) 学会誌・学会報等の発行
- (3) その他本会の趣旨に沿う事業

第4条 本会の会員は一般会員, 学生会員, 団体会員, 賛助会員, 功労会員及び名誉会員よりなる。

- (1) 一般会員は本会の趣旨に賛同する個人とする。
- (2) 学生会員は本会の趣旨に賛同する個人のうち, 大学又は大学院に籍を置く学生とする。
- (3) 団体会員は本会の趣旨に賛同する大学, 研究所, 図書館その他の研究・教育団体とする。
- (4) 賛助会員は本会の趣旨に賛同する企業等とする。
- (5) 功労会員は本会の活動に長く寄与した個人とする。功労会員の規程は別に定める。
- (6) 名誉会員は本会の活動に特別に寄与した個人とする。

第5条 本会の会費について以下の通り定める。

- (1) 会員は所定の会費を納めるものとする。
- (2) 会費の額については次の通りとする。

一般会員	年額	6,000円 (在外会員は年額12,000円)
学生会員	年額	3,000円 (在外会員は年額10,000円)
団体会員	年額	5,000円
賛助会員	年額	15,000円

- (3) 会費は入会時点又は会計年度開始時点で納入する。
- (4) 2年間にわたって会費納入がない場合は会員の資格を失う。
- (5) 名誉会員, 功労会員, 顧問からは会費を徴収しない。

第6条 本会の会計年度は4月1日に始まり、翌年3月31日をもって終わる。

第7条 本会に執行部、事務局、役員会、学会誌編集委員会、学会賞選考委員会、大会実行委員会、研究会（SIG）を置く。

- (1) 執行部は会長、副会長、事務局長、事務局員で構成し、本会全体にかかわる事業を執行・監督する。
- (2) 事務局は事務局長及び事務局員で構成し、本会の事務を執行する。
- (3) 役員会は役員で構成し、本会にかかる諸問題を審議・決定する。
- (4) 学会誌編集委員会は学会誌の刊行にかかる業務を担当する。学会誌編集委員会の規程は別に定める。
- (5) 学会賞選考委員会は学会賞・奨励賞の選考にかかる業務を担当する。学会賞選考委員会の規程は別に定める。
- (6) 大会実行委員会は大会の企画・準備・実施にかかる業務を担当する。大会実行委員会の規程は別に定める。
- (7) 研究会（SIG）は会員のうち、希望する者によって構成し、それぞれが掲げる研究目的に応じた活動を行う。研究会の規程は別に定める。

第8条 本会に次の役員をおく。

- (1) 会 長 1名
- (2) 副会長 若干名
- (3) 理 事 若干名
- (4) 幹 事 若干名

第9条 役員の任期は以下の通りとする。

- (1) 会長・副会長の任期は2年とし、引き続き2期までの再任を妨げない。
- (2) 理事・幹事の任期は2年とし、再任を妨げない。
- (3) 任期は当該年度の4月1日から起算する。
- (4) 役員の定年を70歳とする。任期の途中で定年に達したときは当該年度の終了まで、その任にあたる。

第10条 役員の任務は以下の通りとする。

- (1) 会長は本会を代表し、会務を統括する。会長は総会・役員会を招集し、これを主宰する。
- (2) 副会長は会長の命ずる職務を所掌するとともに、会長を補佐し、必要に応じて会長の職務を代行する。
- (3) 理事は役員会に出席し、本会の運営に関わる重要事項を審議・議決する。
- (4) 幹事は役員会に出席し、理事を補佐し、本会の運営に関わる重要事項を審議議決する。

第11条 役員は役員会における投票によって決定する。

第12条 本会に次の役職員をおく。

- | | |
|-----------------|-----|
| (1) 顧問 | 若干名 |
| (2) 事務局長 | 1名 |
| (3) 事務局員 | 若干名 |
| (4) 監査 | 1名 |
| (5) 学会誌編集委員会委員長 | 1名 |
| (6) 学会誌編集委員 | 若干名 |
| (7) 学会賞選考委員会委員長 | 1名 |
| (8) 学会賞選考委員 | 若干名 |
| (9) 大会実行委員会委員長 | 1名 |
| (10) 大会実行委員 | 若干名 |

第13条 役職員の任期は以下の通りとする。

- (1) 顧問の任期は終身とする。
- (2) 事務局長・事務局員、監査、学会誌編集委員会委員長及び委員、学会賞選考委員会委員長及び委員の任期は2年とし、引き続き2期までの再任を妨げない。任期は当該年度の4月1日から起算する。
- (3) 大会実行委員会委員長及び委員の任期は、役員会で承認された日から当該大会に係る業務の終了時までとする。
- (4) 顧問を除く役職員の定年を70歳とする。任期の途中で定年に達したときは当該年度の終了まで、その任にあたる。

第14条 役職員の任務は以下の通りとする。

- (1) 顧問は役員会の求めに応じて学会運営への助言を行う。
- (2) 事務局長は事務局を主宰し、本会の事務を執行・監督する。
- (3) 事務局員は事務局長の指示の下、必要な業務を執行する。
- (4) 監査は本会の会計及び運営が適切になされているか精査し、その結果を総会で報告する。
- (5) 学会誌編集委員会委員長は学会誌編集委員会を主宰し、学会誌の刊行にかかる業務を執行・監督する。
- (6) 学会誌編集委員は委員長の指示の下、必要な業務を執行する。
- (7) 学会賞選考委員会委員長は学会賞選考委員会を主宰し、学会賞・奨励賞の選考にかかる業務を執行・監督する。
- (8) 学会賞選考委員は委員長の指示の下、必要な業務を執行する。
- (9) 大会実行委員会委員長は大会実行委員会を主宰し、大会の企画・準備・

実施にかかる業務を執行・監督する。

(10) 大会実行委員は委員長の指示の下、必要な業務を執行する。

第15条 役職員は会長が推薦し、役員会で承認する。役職員と役員の兼務を妨げない。

第16条 本会は以下の会議を開催する。

- (1) 総会は会長の招集により、原則として年1回以上開催し、会則の改定、予算・決算その他重要事項を審議する。なお、電子メールやその他の手段を用いた総会の開催も可能とする。総会での議決は出席者の過半数による。
- (2) 役員会は会長の招集により、原則として年2回以上開催し、本会の運営にかかる諸問題を審議し、決定する。なお、電子メールやその他の手段を用いた役員会の開催も可能とする。役員会での議決は出席者の過半数による。
- (3) 事務局会議は事務局長の判断の下、不定期に開催する。
- (4) 学会誌編集委員会、学会賞選考委員会、大会実行委員会は各委員長の判断の下、不定期に開催する。

付則 本会則は2020年4月1日から施行する。

英語コーパス研究（第27号）

【2020年10月31日発行】

編集・発行 ©2020 英語コーパス学会

〒819-0395 福岡県福岡市西区元岡744

九州大学大学院言語文化研究院 内田諭研究室気付

E-mail (事務局長) : jaecs.hq@gmail.com

Twitter: @JAECS2012

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

郵便振替口座：009300-3-195373（英語コーパス学会）

印刷所 株式会社ニシキブリント

〒739-2117 東広島市高屋台2-1-12

English Corpus Studies: Vol.27 2020

Articles

Yasunori NISHINA / Toward Improving Lexical Descriptions in a Japanese-English Dictionary based on the Japanese-English Parallel Corpus WikipediaKyoto-LWP: The Case of <i>Katameru</i>	1
Tatsuya ISHII and Takeshi KAWAMOTO / The Behavior of Adverbs in the Results Sections of Experimental Medical Research Articles: A Corpus-Based Move Analysis	23
Michiko YAGUCHI / The Process of Grammaticalization of <i>be bound to</i> as a Semi-Auxiliary in American English.....	53

Note

Satoshi YAMAZAKI / The Expansion of the <i>how about</i> Construction in American English	73
--	----

Conference Program & Abstract

The 45th Conference of Japan Association for English Corpus Studies	87
---	----