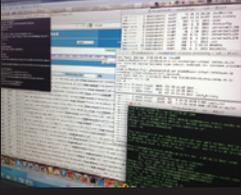


JAECS
Japan Association for English Corpus Studies

ISSN 1340-301 X

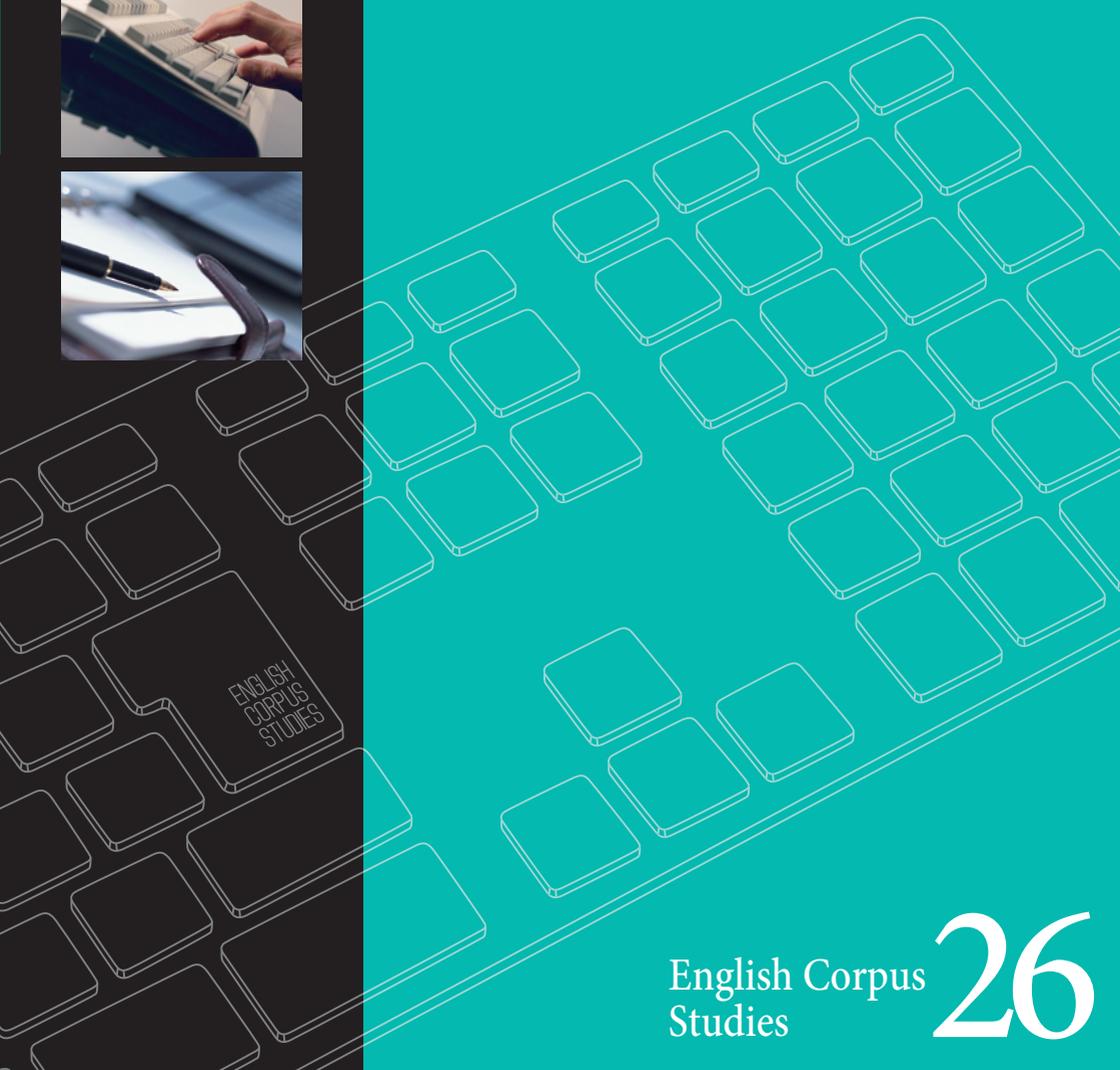
2019

英語コーパス学会



英語コーパス研究

第26号



English Corpus
Studies

26

2018年度 英語コーパス学会役員

会 長：投野由紀夫

副 会 長：井上 永幸

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, 井上 永幸,
梅咲 敦子, 岡田 毅, 金澤 俊吾, 小島ますみ,
高橋 薫, 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：加野まきみ

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>新井 洋一, 五百蔵高浩, 石川 保茂, 高橋 薫,
中條 清美

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 26 号

英語コーパス学会

2019

目 次

論文

Dickens の *The Mystery of Edwin Drood* と T. P. James によるその続編の文体的評価
— 多変量解析を用いて —
..... 後藤 克己 1

英語における live a/an Adj life と lead a/an Adj life に見られる
意味的違いについて
..... 金澤 俊吾 21

The Time Span Required for Syntactic Changes in the History of English
..... Satoru TSUKAMOTO 39

資料

英語コーパス学会第44回大会資料63

「論文」

Dickens の *The Mystery of Edwin Drood* と T. P. James によるその続編の文体的評価 — 多変量解析を用いて —

後藤 克己

Abstract

Three years after Dickens' death, Thomas Power James (henceforth James) added a continuation to *The Mystery of Edwin Drood*, claiming that it was written by the 'spirit-pen of Charles Dickens, through a medium'. This study attempts to clarify whether the continuation can be considered a posthumous work of Dickens as James suggested, at least as far as the linguistic features are concerned. Word preferences in the continuation are analyzed for similarity with those in *The Mystery of Edwin Drood*. Methods used are multivariate analyses of the frequencies of frequent words in three corpora - *The Mystery of Edwin Drood*, the continuation, and *Our Mutual Friend*, with the third corpus added as a reference. The analyses display distinct clustering of sections included in the continuation, on one hand, and those in the two Dickens' works, on the other, highlighting differences in terms of word preferences. In addition to the findings that the continuation is stylistically distinct from the original so that James' claim is not supported, this paper also confirms that (i) the most distinct results are obtained by word-list of narrative - not speech or combinations of both, and that (ii) high frequency words, up to 100, sufficiently distinguish the authors.

1. はじめに

1.1 リサーチクエスション (RQ)

Charles Dickens は *The Mystery of Edwin Drood* (以下, *ED*) を Serial Fiction として 1870 年 4 月から逐次発表していたが, 作品を完成させることなく同年 6 月に世を去った。その 3 年後, 米国 Vermont 州 Brattleboro の印刷工 Thomas Power James (以下, James) は, *ED* に第 2 部として 23 章もの大部となる続編

を加え完全版として発表した(Dickens [James], 1873)。Jamesはこの続編を, “By the Spirit Pen of Charles Dickens, through a Medium.”, つまり「Jamesが, 亡きDickensの霊に導かれてテキスト化したもの」とアピールしている。本研究では, この続編をめぐる次の3点について考察する。

(RQ1) テキストに生起する語彙頻度を多変量解析の手法を用いて分析することで, EDと続編の文体の異同性を語彙嗜好の観点から評価し, 続編の文体がDickensの霊に導かれたものと考えうるほどに, Dickensのそれと類似しているかどうかを判定する。

(RQ2) 頻出語彙数の多変量解析による著者推定では, 多くの場合テキストから発話部を除き, 分析対象を地の文に生起する語彙に限定している(Hoover, 2001, 2002他)が, 発話テキストの除去は, コーパスによっては1重引用符とアポストロフィの区別, 引用符を欠いたFree Direct/Indirect Discourseの抽出などが必要となり, 特に分析対象コーパスが大規模な場合大きな負担となる。そこで発話部を除外せずテキスト全体の語彙で分析した場合と, 地の文のみの語彙を用いた場合の結果を比較し, 分析対象を地の文に限ることの優位性について分析の精度の観点から評価する。

(RQ3) 頻出語彙数の多変量解析による著者推定で用いる語彙の数(タイプ数)については, 推定方法は異なるものの, 上位150語(Burrows, 2002: pp. 276-67), あるいは同800語(Hoover, 2004: p. 455)で最良の結果が得られたとの研究がある。一方で, 複数作品をそれぞれ複数セクションに分割してクラスター分析した場合, 上位10語でも各セクションがそれぞれの作者に正しくグループ化されるとの報告もある(Hoover, 2003b: p. 343)。そこで, 語彙頻度データが, その頻度ランクの高低によって分析結果にどう影響するかを評価し, 著者推定において十分と考えられる語彙の数を見究める。

1.2 研究の背景

1.2.1 Jamesの続編への批評

Jamesによる続編(以下, 続編)については, Jamesが「Dickensの霊による作¹⁾」(Dickens [James], 1873, 中表紙)とアピールしたこともあり, 物語性, 人物造型の一貫性, 言語的側面等の観点から以下に概括するように多くの批評がなされた。

B., W. H.²⁾ (1874: p. 219, pp. 221-23)は, 続編を「Dickens風の文体の癖をかるうじて真似ていて, そこここに注意深くしつらえた感傷を, しかるべく添えてはいるが, 言葉づかいはウィットに欠け, むだ口が目につく」と評している。

また、主要な登場人物 (Mr. Grewgious, Miss Twinkleton, Rosa, Jasper など) が、一見 ED と似たキャラクターとして描かれているものの、例えば「Grewgious の発話からはユーモアが失われて陰気／単調なものとなり、Rosa のそれからは面白さが消えて単に愚かさが目立ち、Mr. Crisparkle の発話からは快活さが消え重々しい常套句に終始している」として、人物造型の一貫性の欠如を指摘している。さらに言語的側面について、文法面で自動詞と他動詞の誤用 (例, lay), 主語・動詞の数の不一致, 不規則動詞の過去変化形の誤用などを挙げ、さらに、realise [realize]³ を perceive の意味で用いていること, happen でなく transpire を, as soon as でなく directly を使用していることなど、多くの Yankeeisms (アメリカ英語風の語彙使用) を指摘している。

また、Gadd (1905: p. 272) は、続編について、「まるで原典 [ED] のもつ非凡さをうかがわせるものがない」と批評し、また人物造型の不自然さを指摘するとともに、言語的側面では「気紛れとも言うべきひどい文法乱れ」を指摘しているが、その内容について具体的に言及していない。

次に、Walters (1913: p. 217) は続編を、「無価値であり、書き間違い・非慣用語法が目立ち、ひどくアメリカ的な型にはまった作」と Gadd (1905) による関連言説を引用して否定的に評している。ただし、「書き間違い・非慣用語法、ひどくアメリカ的型はまり」について具体的に例示していない。

そして「シャーロック・ホームズ」シリーズの著者であり、また、心靈現象の観点から Dickens, Oscar Wilde, および Jack London の遺作を研究した Doyle (1927: pp. 343-44, p. 349) は、批評は交霊云々ではなく物語自体についてなされるべきとして、「似てはいるが退屈な Dickens 作品になっている。わくわく感、ひらめき、伸び伸び感がない。しかし趣向や文章上の癖に Dickens らしさが残っている」と評している。また、原典と同様に語りが突然現在形になること、登場人物にグロテスクなニックネームをつけ、それを大文字で綴っていることを指摘するとともに、“traveller” のようにエル 2 つで綴られていることを例にあげて綴りがアメリカ英語風でない等々と述べている⁴。そして続編を、Dickens についての知識がなく、また新聞の小記事さえ書いたことがないと伝えられる米国人 [James] が著したということは、深いトランス状態にあったことの証左だとして、「Dickens の霊による作」とアピールされた続編に一部好意的な批評⁵ をしたうえで、全体として「[続編は] 決定版とは言い難いものの、丁寧に考察するに価するものだ」と批評を締めくくっている。

更に続編発表の1世紀後、Wolkomir (1973) は当時の新聞記事、前述した Doyle の著作、さらに Vermont の史家である Hill (1961: pp. 178-82) の著作から、「Dickens の霊による作」にかかわる記述を引用して肯定的に評している。なお、Cox (1998: p. 569) はこの評について、掲載された *Psychic* ならではのものとコメントしている。

1.2.2 頻出語彙の生起数による著者推定に用いられる異なり語の数

Burrows (2002: pp. 275-77) は独自の手法“Delta”⁶を用いて、英国復古期(17C 後半)の25人の作者による韻文を参照コーパスとし、32作の長詩について、頻度上位40～150語で著者推定を行った。その結果、上位150語を用いた場合に最もよい結果が得られたとしている。

Hoover (2004: p. 456) は、その“Delta”を散文に適用して、20人の作者による米国小説(19C末～20C初頭)を参照コーパスとし、39作について上位20～800語で著者推定した。その結果、推定の確度は語彙数の増加につれて高くなり、700語で最もよい結果が得られたとしている。また Hoover (2001, 2002, 2003a) は、「14人の作者による29作」のような多くの作者・作品で構成されたコーパスに頻出する語、語シーケンス、および語コロケーションの頻度をクラスター分析して行う著者推定では、上位600～800語で最良の結果が得られたとしている。

一方、Hoover (2002: p. 157) は Burrows (1987) を引用して、全トークン数のほぼ20%を占める *the, and, of, a* および *to* の語は、直観的には意味的・文体的に重要でないように思えるが、作者、作品さらには1つの作品内で登場人物さえ互いに区別する、と述べている。さらに Hoover (2003b: p. 343) は、そのメインテーマである同一作品内における文体差異を、上位50～800語で分析する前段として、クラスター分析による著者推定の有効性を評価する中で、作者の異なる6作品の地の文を、それぞれ25,000語のセクションに分割して、各セクションに頻出する語彙の生起数をクラスター分析した場合、上位10語でも各セクションをそれぞれの作者に正しくグループ化していることに言及している。

2. 分析方法

本研究では、分析対象作品の語彙嗜好に類似性があるかどうかを明らかにするため、それらの作品に生起する語彙の頻度を多変量解析の手法を用いて分析する。

2.1 分析ツール

続編が James の手によるものであることは疑いない。本研究の RQ1 は、続編の文体が James の主張するように Dickens の霊に導かれてテキスト化されたもの、と考えるほどに類似しているか否かである。一般に著者推定は、候補となる著者の作品を特徴づける言語データを、問題作品の特徴と比較して行われる。Dickens の作品は多く遺されていて、その文体の特徴を表すデータ（教師データ）は容易に得られるが、James の手による作品は、ごく短い断片を除き他に見つかっていない⁷。従って作品をカテゴリ化する教師あり学習（supervised learning）の手法はとれない。しかし、RQ1 は続編の文体が Dickens のそれと似ているかどうかの問題であることから、似た者同士でグループ化する手法、具体的には多次元尺度法（multidimensional scaling (MDS)）と階層的クラスタ分析（以下、クラスタ分析）を用いた⁸。これらは先行研究でも用いられている（田畑, 2016; Hoover, 2001, 2002 他）。

2.2 分析データ（コーパス）

分析作品は *ED* と続編の 2 作に、同じく Dickens による作品で *ED* より数年前の作品である *Our Mutual Friend* (1864-65)（以下、*OMF*）を参照用に加えた計 3 作とした⁹。*ED* と続編については、Dickens [& James] (1873)¹⁰ を自ら OCR スキャンして得た電子テキストを元の書籍と目視照合し、誤変換された文字を修正してコーパス化した。また *OMF* のテキストについては、Project Gutenberg からダウンロードした電子テキスト¹¹ を使用した。

これらを地の文と発話部に分離¹² し、分析結果のプロットや樹形図で、語彙嗜好が作品間で類似しているかどうかを把握しやすくするため、それぞれその規模に応じて概ね 1 セクションが 15,000 トークンとなるように章単位で分割した（*ED* と続編はそれぞれ 4 セクション、*OMF* は 12 セクションの計 20 セクション）。また、各セクションの語彙嗜好を際立たせるため、語彙の生起数は表記形でなくレマ¹³ でカウントした。これら 3 作の地の文・発話部別のトークン数とタイプ数（レマ）を表 1 に示す。

表 1 3 作品の規模データ

作品区分	地の文			発話部		
	<i>ED</i>	続編	<i>OMF</i>	<i>ED</i>	続編	<i>OMF</i>
トークン数	54,115	71,826	174,311	49,511	57,403	159,471
同、比率	52%	56%	52%	48%	44%	48%
タイプ数	5,440	4,041	8,744	3,547	3,416	6,017

なお、語彙の内、人物・場所などの固有名詞や、作品テーマに関連する語彙は、作者の文体とは関係なく、むしろその異同性の分析結果をゆがめることになりかねないので除外する必要があるが、手作業でこのような語彙を選別するのは容易ではなく、また選別に際して恣意的な判断が混入する懸念が残る(田畑, 2016: p. 60)。そこで本研究では、Hoover (2004) を参考に、「全体の語数の70%以上が1作品に偏在する語彙を除外する」手法を用い、人物名についてはさらに手作業で補完的に除外した。なお、続編にはその性格からEDと同じ固有名詞が含まれるため、この「70%の偏在基準」を、(i) ED+ 続編, OMFの2作品, (ii) ED, 続編, OMFの3作品, の2ステップで適用した。また地の文に生起する人称代名詞の頻度は、語彙嗜好とは無関係な語りの人称の違いによる影響を受ける(Hoover, 2001: pp. 427-28) ため除外した¹⁴。これらの語彙選別を経たテキスト全体・地の文・発話の3種類の語彙頻度データを、語彙(行)と20のセクション(列)で表にし、各語彙の作品ごとの相対頻度の計で降順にソートし、上位50, 100, 200, 300, 500, 750, および1,000の7つの語頻度ランクで切り出して分析に用いる語彙データとした(表2, 地の文の例)。

表2 語彙頻度データの概略構造

		ED 4セクション				続編 4セクション				OMF 12セクション			
		O1N	O2N	...	O4N	P1N	P2N	...	P4N	M1N	M2N	...	MCN
語頻度 (レマ)	the	786	826	...	759	847	1221	...	933	870	1032	...	713
	and	456	514	...	507	432	721	...	493	491	713	...	505
	be	347	371	...	358	420	640	...	504	378	517	...	396
	of	436	464	...	382	301	451	...	363	433	606	...	375
	to	271	336	...	351	410	645	...	479	324	433	...	326
	a	335	329	...	330	281	445	...	302	421	472	...	323
	in	260	274	...	247	201	319	...	265	309	370	...	278
	have	115	204	...	207	240	326	...	280	165	258	...	204
	that	111	149	...	169	237	358	...	309	139	220	...	202
	with	161	183	...	197	122	191	...	150	184	280	...	211
	mr	134	155	...	151	146	203	...	148	137	202	...	138
	as	95	153	...	140	62	83	...	37	133	188	...	225
	say	121	197	...	237	89	172	...	100	142	283	...	166
	at	73	110	...	110	106	158	...	114	125	172	...	126
	on	109	89	...	115	98	116	...	66	112	134	...	97
	for	75	90	...	83	109	186	...	119	65	102	...	76
	s	54	62	...	99	76	139	...	95	64	124	...	95
	by	73	85	...	59	56	95	...	53	87	102	...	70
	but	37	72	...	66	66	96	...	68	60	101	...	63
	not	28	50	...	60	75	132	...	80	46	65	...	47
...	
...	

3. 分析結果

3.1 3つの作品における語彙嗜好の異同性

3つの作品について、地の文・テキスト全体・発話の3種類のデータの上位100語をMDSで分析した結果を図1に示す。ここで、プロットのセクションラベルは以下のとおりである。(以下、同)

- (i) 先頭は作品の区別で、‘O’はED (Original part), ‘P’は続編 (Posthumous part), ‘M’はOMF (Mutual)
- (ii) 2番目の1,2,・・・,9,A,B,Cは作品内のセクションの区別
- (iii) 3番目は語彙データの種別で、‘N’は地の文, ‘T’はテキスト全体, ‘S’は発話
また図下部のハイフンで連ねた2つの比率値は、それぞれ1軸－2軸の寄与率である。

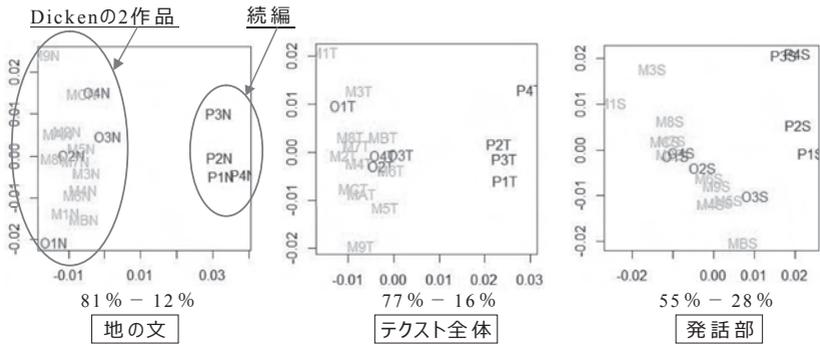


図1 上位100語によるMDSプロット

つぎに同様に上位100語についてクラスター分析して得た樹形図を図2に示す。ここで図下部のハイフンで連ねた2つの数値は最終段階のクラスター対の距離 (clustering height¹⁵; 以下、クラスター高), およびそれとその一つ前のクラスター高の比 (最後のクラスター高 / (最後 - 1) のクラスター高; 以下、ハイト比) である。

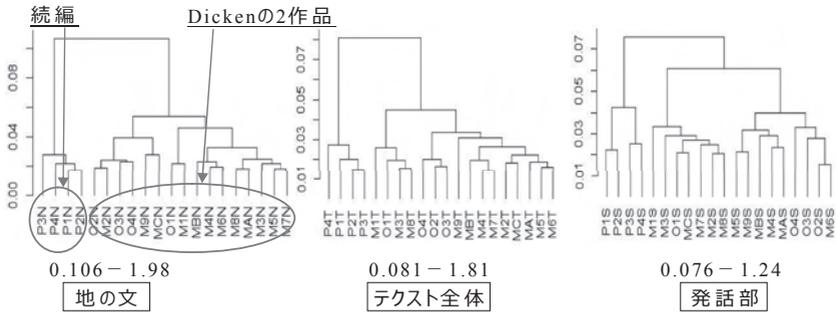


図2 上位100語による樹形図

図1・図2の各左端の地の文語彙によるプロット／樹形図には、Dickensの2作品のセクションは1つのクラスターとなり、続編のセクションはそれとは別クラスターを形成することが明瞭に表れている。また、最後のクラスター高(0.106)、およびハイト比(1.98)が、この3種類の分析の中で最も大きく、地の文語彙による分析で2つのクラスターの区別性が最も高くなることを示している。

また、同2図中央の、テキスト全体語彙によるプロットでは、地の文の場合より、やや各セクションが拡散し、また、樹形図では最後のクラスター高が24%程度狭まっている(0.081)。しかし、ハイト比は1.81で9%の減少に留まっていて、依然として2つのクラスターの区別性が高いことがうかがえる。

ちなみに、同2図右の、発話部語彙によるプロット／樹形図ではDickens 2作品と続編は別クラスターを形成しているものの、プロットでは各セクションがテキスト全体の場合より更に拡散している。樹形図で最後のクラスター高は28%程度の減少に留まっている(0.076)ものの、ハイト比は1.24へと大きく減少し、2つのクラスターの区別性が低下していることがうかがえる。ここで、ハイト比が1に近いことは、クラスタリングの最終段階におけるクラスター対の隔たり(ここでは、Dickensの2作品と続編の間)が、その1つ前のクラスタリングにおけるクラスター間の隔たり(ここではDickensの2作品のセクションによるクラスター間)と大差ないこと、つまりDickensの2作品と続編の語彙嗜好の差異が、Dickensの2作品の中での語彙嗜好の差異と比較して抜きん出たものではないことを示している。

同様に、上位500語について分析した結果を図3と図4に示す。

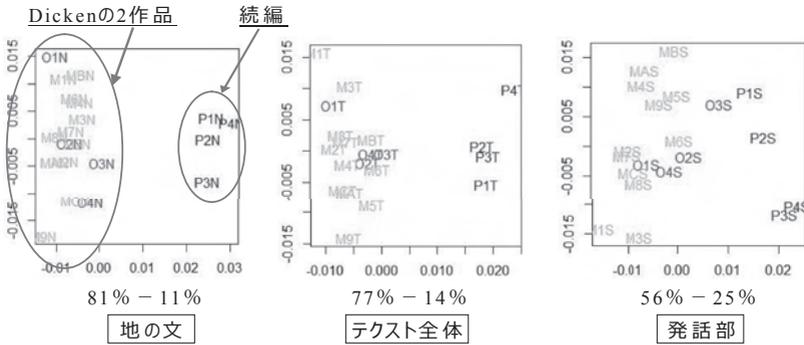


図3 上位500語によるMDSプロット

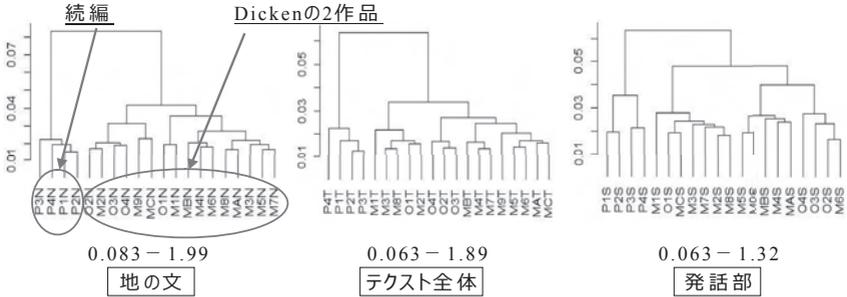


図4 上位500語による樹形図

上位 500 語による分析においても、上位 100 語の場合と同様な結果と傾向が 3 種類のデータに見られる。語彙数が増加したことで Dickens の 2 作品と続編のクラスター間の距離は 100 語の場合の概ね 80% に減少¹⁶しているものの、クラスターの区別性の尺度になると考えられるハイト比は殆ど同じである。これら 2 つの語彙ランク以外による分析でも同様であり、上位 50 語～1,000 語の 7 つの語彙ランクのデータで一貫して、Dickens の 2 作品のセクションは一つのクラスターを形成し、続編のセクションはそれとは別クラスターを形成していた。

さらに地の文・テキスト全体・発話の 3 つの語彙データを一括した、上位 500 語のデータ¹⁷による MDS のプロットにおいても、Dickens の 2 作品と続編のセクションはそれぞれ別クラスターを形成し、それらの距離は地の文、テキスト全体、発話の順に狭まっていた (図 5)。

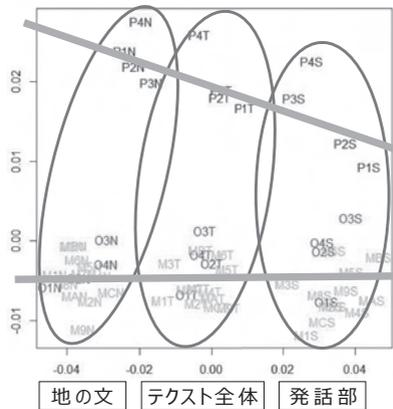


図5 上位500語の3種類の語彙データを一括したデータによるMDSプロット

3.2 頻度ランクによる語彙嗜好の違いと分析結果への影響度

前3.1節の上位50語～1,000語の7つの語彙データによる分析結果には、一貫してDickensの2作品と続編の間の語彙嗜好の違いが表われている。7つの頻度ランクについて、地の文の語彙データをクラスター分析して得た樹形図の、クラスター高とハイト比を比較した(表3, 図6, 図7)。

表3 頻度ランクとクラスター高・ハイト比

語彙頻度ランク	クラスター高		ハイト比
	last	last-1	
1-50	0.121	0.063	1.94
1-100	0.106	0.054	1.98
1-200	0.094	0.048	1.97
1-300	0.089	0.044	1.99
1-500	0.083	0.042	1.99
1-750	0.080	0.040	1.99
1-1000	0.078	0.039	1.99

凡例

last: 最終段階のクラスタリングにおけるクラスター高

last-1: 最終段階の一つ前のクラスタリングにおけるクラスター高

ハイト比: lastとlast-1のクラスター高の比 (last/ (last-1))

(以下, 同)

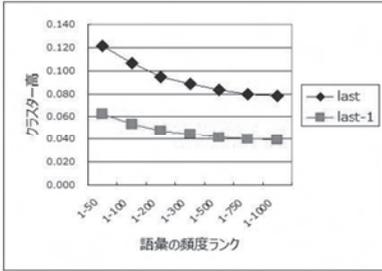


図6 頻度ランクとクラスター高

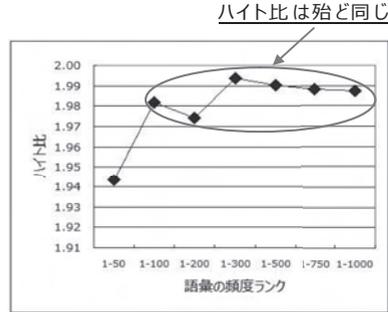
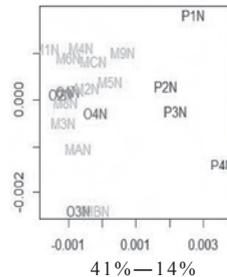
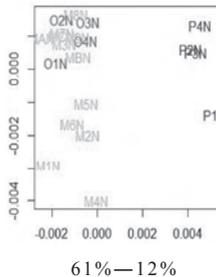
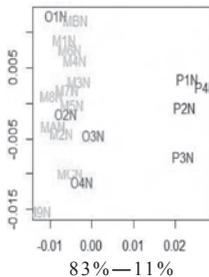


図7 頻度ランクとハイト比

分析語数が増えるにつれクラスター高が減少している¹⁸ (図6) が、ハイト比は100語と1,000語の間でごく僅かな差しかなく、約2倍の比を保っている(図7)。つまり語彙数が100語と1,000語の間で大きく変わっても、分析結果に大きな差は表れていない。

そこで、3.1節の結果からDickensの2作品と続編の語彙嗜好は異なっているとした上で、この1,000語のそれぞれの頻度ランクの語彙データが、どのようにこの結果に反映されたかを分析した。すなわち、前節の分析結果から区分性が最も高いと考えられる地の文について、語彙の属するランクレベル(つまり、生起数の多い高ランク語彙、少ない低ランク語彙など)による語彙嗜好の違い、および分析結果への影響度を確認することとし、上位1,000語までの語彙を100語ごとにセグメント化したデータを用いて前節と同様に分析した¹⁹(図8, 図9)。なお、以下の図で、プロット/樹形図の下に付記したデータの内容は、前節と同じである。



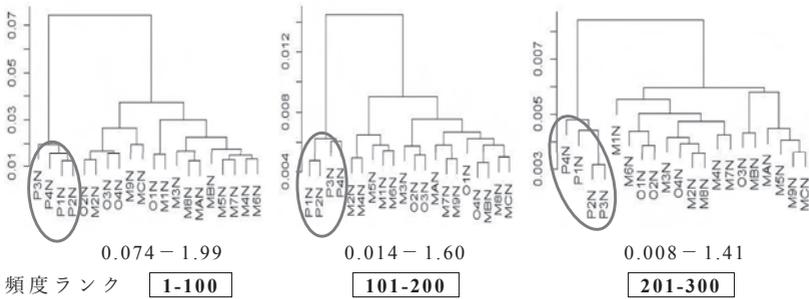


図8 地の文語彙の上位300語を100語ごとにセグメント化したデータによる分析結果

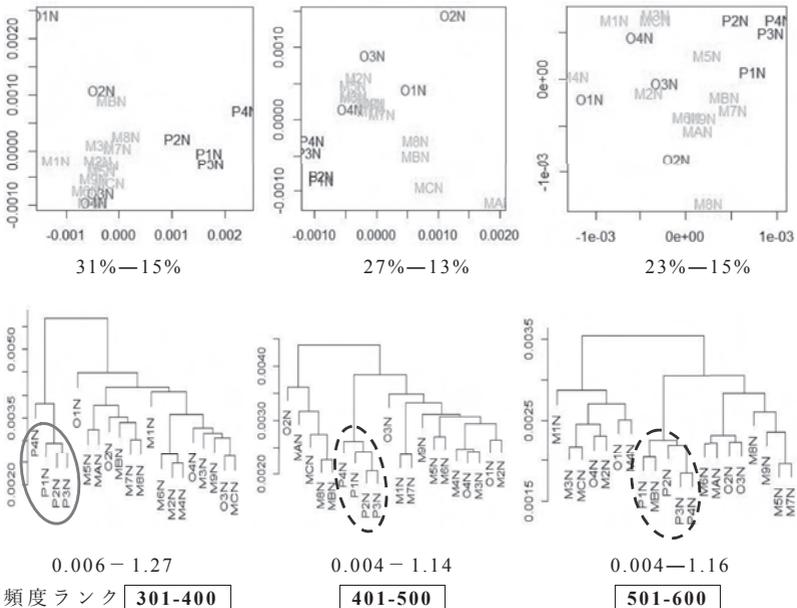


図9 地の文語彙の上位301~600語を100語ごとにセグメント化したデータによる分析結果

上位400語までの4つの100語セグメントによる分析では、前節の上位50語~1,000語による分析で一貫して得られたのと同様に、Dickensの2作品と続編のセクションが最終段階のクラスタリングで別クラスターとなり、続編のセクションが単独のクラスターとしてグループ化されている。一方、401語以降の100語セグメントでは、続編のセクションは最早最終段階の単独クラス

ターになっていない。(601 語以降の 100 語セグメントでも同様であった。) 最後のクラスター高は、ランク 1-100 では最大の 0.074 であったが、次のランク 101-200 では 0.014 と約 1/5 に大きく減少している。また、ハイト比は、ランク 1-100 の最大の 1.99 から、ランク 101-200 以降の 1.60, 1.41, … と徐々に減少している。

一方、続編のセクションが最後の単独クラスターとしてグループ化されなかったランク 401 以降の 100 語セグメントでは、最後のクラスター高はごく小さく、またハイト比は 1 に近い。従ってランク 401 以降の語のクラスタリングへの影響度はごく低いものと考えられる。

本研究ではセクション間の距離をユークリッド法で測っている。セクション間の語彙頻度のバラつきは一般に高頻度語ほど大きく、またセクション間の距離は、各語彙の頻度差の 2 乗値の計の平方根であるため、一般に高頻度語ほど距離への影響度が高い。さらに、作品には少数の頻度上位語が際立って多く生起する傾向があるため、クラスター高は高頻度語のデータをより反映したものとなる。本節の分析結果もこれと符合しており、分析結果に大きく影響するのは、ランク 1-100 のセグメントの語頻度データであって、ランク 101 以降のセグメントに属する語データによる影響は低いと考えられる (表 4, 図 10, 図 11)。

表 4 100語セグメントのクラスター高とハイト比

語彙頻度ランク	クラスター高		ハイト比	続編の区分性
	last	last-1		
1-100	0.074	0.037	1.99	最終クラスタリングでDickensの2作品と続編を区分
101-200	0.014	0.0091	1.60	
201-300	0.0084	0.0060	1.41	
301-400	0.0057	0.0045	1.27	続編セクションは最終の単独クラスターにならない
401-500	0.0044	0.0038	1.14	
501-600	0.0035	0.0031	1.16	
601-700	0.0027	0.0025	1.06	
701-800	0.0030	0.0024	1.22	
801-900	0.0023	0.0022	1.07	
901-1000	0.0021	0.0020	1.03	

101-200 のセグメントでは大きく減少している

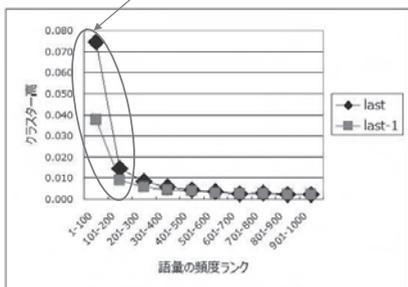


図10 100語セグメントのクラスター高

ランク 401 以降では 1 に近い

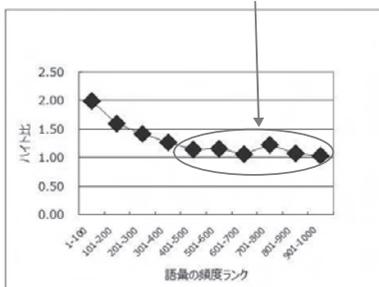


図11 100語セグメントのハイト比

次に、頻度ランク上位 25 語について分析した (図 12 ~ 図 14)。

上位 25 語以上ではクラスター高・ハイト比は高く、安定している

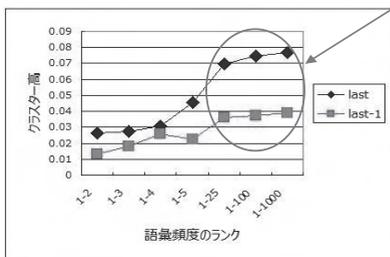


図12 頻度ランクとクラスター高 (上位25語)

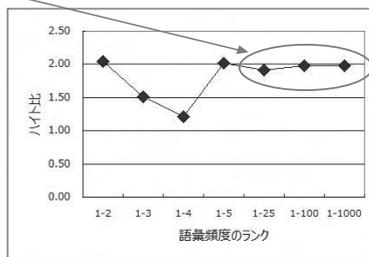


図13 頻度ランクとハイト比 (上位25語)

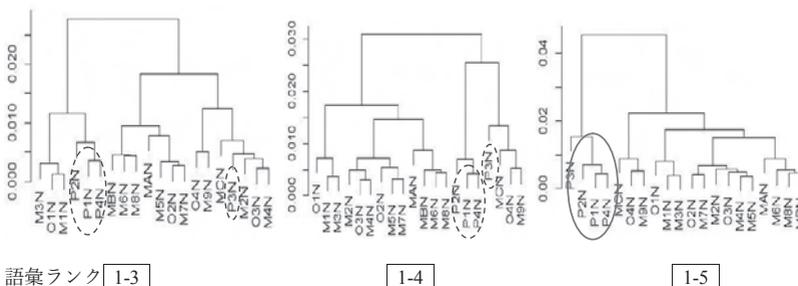


図14 上位 5 語までの樹形図

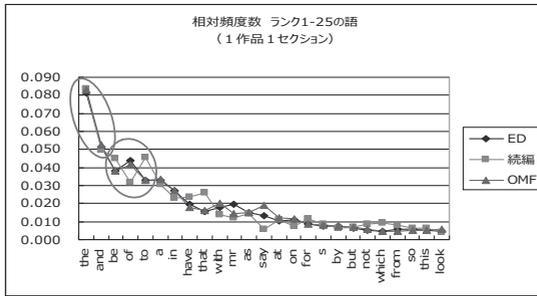


図15 上位25語の相対頻度数

一般にクラスタリングに大きく影響するのは、高頻度かつセクション間で頻度差が大きい語である。本研究で扱った3作品の場合、図15のように上位の2語 (*the*, *and*) は際立って高頻度だが、ランク3以降の *be*, *of*, *to* とは異なり、作品間の頻度差が小さい。そのためランク4までの語彙による樹形図では、続編セクションが最終の単独クラスターとなっていない (図14)。一方、ランク1-25以降では2つのクラスター高およびハイト比は安定しており (図12, 図13), また Dickens の2作品と続編の区分がMDSプロットおよび樹形図に明瞭に表れていることから、上位25語程度が絞り込みの限界と考えられる。なお、同サイズのテキストで、*the* と *and* の生起数が2倍異なることもある (Hoover, 2002: pp. 157-58) ことから、一般に分析語数を上位5語程度の少数にまで絞り込むことは適当でないと考えられる。

前3.1節および本節では、データの距離・結合法に、それぞれテキストマイニングで多く用いられるユークリッド距離・ワード法を用いている。比較のため、使用した3作品のコーパスの上位100語と500語について、*dist* 関数の3つの距離 (euclidean, manhattan, canberra) と *hclust* 関数の4つの結合法 (ward, D2, single, complete, average) を組み合わせで分析した (表5, 表6)。

表5 上位100語による比較

結合法 (hclust)	距離 (dist)	クラスター高		ハイト比
		last	(last-1)	
ward.D2 ワード法	euc	0.106	0.054	1.98
	man	0.654	0.278	2.35
	can	45.0	18.8	2.39
single 最近隣法	euc	0.039	0.028	1.38
	man	0.238	0.162	1.47
	can	17.7	13.7	1.29
complete 最遠隣法	euc	0.065	0.050	1.28
	man	0.368	0.265	1.39
	can	26.2	19.5	1.34
average 群平均法	euc	0.048	0.037	1.30
	man	0.298	0.202	1.47
	can	21.8	15.9	1.37

備考：euc: euclidean, man: manhattan, can: canberra

表6 上位500語による比較

結合法 (hclust)	距離 (dist)	クラスター高		ハイト比
		last	(last-1)	
ward.D2 ワード法	euc	0.083	0.042	1.99
	man	0.798	0.365	2.19
	can	292	182	1.61
single 最近隣法	euc	0.031	0.023	1.37
	man	0.340	0.261	1.30
	can	162	159	1.02
complete 最遠隣法	euc	0.051	0.038	1.33
	man	0.468	0.345	1.35
	can	205	176	1.17
average 群平均法	euc	0.039	0.030	1.30
	man	0.393	0.297	1.33
	can	182	169	1.08

その結果、ワード法の数値が最も大きくなっているが、続編とDickensの2作品は、全ての距離・結合法の組み合わせで一貫して最終のクラスター対を形成しており、他の組み合わせを選んでも分析結果に影響を与えない。

4. 結論

RQ1については3.1節の分析で、最終のクラスタリングにおいて一貫してDickensの2作品のセクションは1つのクラスターを形成し、続編のセクションはそれとは別クラスターを形成した。このことはDickensの2作品と続編の語彙嗜好に明らかな差異があることを示しており、続編の文体がDickensのそれとは異なっていることが裏付けられた。

RQ2については3.1節の分析で、地の文語彙を用いた場合にテキスト全体語彙を用いた場合より明確な嗜好の差異が表れており、著者推定における地の文語彙の優位性をあらためて確認した。なお、テキスト全体語彙で得られる区分性は、地の文語彙の場合より低いものの差は小さく、分析対象となる著者/作品が限定的なケース、あるいは予備的/概略の区分を行う場合であれば、全体語彙を分析に充てることも可能と考えられる。

RQ3については、3.1節でみられた上位50語～1,000語による分析で一貫して得られたクラスタリングパターンが、3.2節の分析で、概ね上位100語あるいはそれ以下の語で形成されており、頻度ランクが下がるにつれパターン形成への影響度が大きく減少することが確認された。これは、MDSやクラスター分析を用いて行う著者推定は、頻度上位100語程度の分析で十分であることを

示唆している。なお前節の分析では上位 25 語でも上位 100 語と同様な結果が得られているが、頻度ランク 4 までの語による分析では結果が異なっている。研究のケースによって分析の対象となる作者・作品は異なり、上位 100 語に含まれる語も異なりうること、また一般に語彙数を減らすことが分析の容易さにつながるわけでもないことから、100 語以下に絞り込むメリットは少ないと考えられる。なお、表 7 は上位 100 語までの語であるが、いずれも平易な日常語と考えられる。

表 7 頻度ランク 1～100の語彙一覧²⁰

ランク 1-25 の語

the, and, be, of, to, a, in, have, that, with, mr, as, say, at, on, for, s, by, but, not, which, from, so, this, look,

ランク 26-100 の語

an, hand, when, upon, do, would, all, out, if, then, take, or, come, one, into, who, little, there, face, man, go, miss, no, time, very, again, eye, some, make, up, more, old, return, see, after, other, down, could, head, reply, now, door, before, turn, what, though, way, leave, think, about, than, house, ask, such, young, two, back, room, know, day, find, seem, over, might, like, gentleman, much, answer, while, place, any, still, well, great, lady

上位 100 語に含まれる語のうち、頻度ランク 21 の *which* は、その殆どが関係詞であり、続編に *ED* の 2.2 倍の頻度（相対頻度、以下同）で生起する。また、主に伝達動詞として用いられる *say*, *return*, *reply* および *ask* のうち、頻度ランク 13 の *say* は 7:3 で *ED* に多く生起し、同 65 の *reply* は 3:7 で逆に続編に多く生起する。（同 58 の *return* と同 78 の *ask* は、それぞれ *ED* と続編にほぼ同じ頻度で生起）。このような関係詞や伝達動詞の選択傾向が、*ED* と続編の文体的差異の例として考えられる²¹。

本研究で用いた、語の生起頻度をクラスター分析（ユークリッド距離・ワード法を適用）する手法は、習慣的に用いられ、同じ作者の作品間で生起頻度が一貫すると想定される高頻度語（Hoover, 2003b: p. 341）が、分析結果に大きく寄与することから、著者推定に有効な手法と考えられる。一方、作者の文体に特徴的でも、生起頻度が低い語の寄与度は小さいため、同じ作者の作品間や、作品内における文体変異の分析のような、微妙な差異に注目する分析への適用には限界があると考えられる。そのような分析に有効な手法の見極めについては、今後の課題としたい。

謝 辞

本稿の内容は、英語コーパス学会第43回大会（2017年9月、関西学院大学）、同44回大会（2018年10月、東京理科大学）、およびTHE 4th ASIA PACIFIC CORPUS LINGUISTICS CONFERENCE (APCLC2018)（2018年9月、高松）における研究発表に基づいている。ご指導いただいた大門正幸教授に感謝致します。また、匿名の査読者の方々から貴重なご指摘・ご助言をいただいた。ここに記して感謝します。

注

1. 以下、日本語訳は全て筆者による。
2. 著者のフルネームは不詳。当批評はW. H. B. と署名されている。
3. W. H. B. の指摘の中ではrealiseと表記されているが、Dickens [James] (1873) ではrealizeの綴りになっている。
4. Doyleは続編全体ではなく、続編の一部を基に批評している。そしてその点について、「Waltersが指摘しているfutilityについて批評するには続編全体を精査する必要があるものの、illiteracyとAmericanismsについては一部による精査で足りる」と述べている(Doyle, 1927: p. 344)。
5. 続編についてDoyleは交霊による作との見方をとるものの、その霊をDickensとすることには懐疑的な立場をとっている。
6. 著者が不明な作品と既知な作品（候補となる参照コーパス）について、高頻度で生起する語の生起特性（z-score）の差を求め、その差の語全体での計“Delta”が最も小さくなる著者を、不明作品の著者と推定する手法である。この手法は、著者候補の手がかりとなる外的証拠が乏しい中で、著者を多くの候補の中から絞り込む研究において有効なツールになりうる(Burrows, 2002: p. 267; Hoover, 2004: p. 454)とされている。本研究では、文体の異同性評価の候補著者はDickensに限定されるため用いなかった。
7. 続編の他に、Dickens [James] (1873) の序文で言及している*The Life and Adventures of Bockley Wickleheap*の断片テキストが見つかったとされているが、続編の規模の0.5%にも満たない600語程度のごく小規模の断片であり、Jamesによる作品の文体特徴を参照するコーパスとして扱えないと考えた。なお、Jamesは当作品もDickensの霊に導かれたものと主張している。
8. すべての分析は、R (version 3.4.4) の環境にてdist, cmdscale およびhclustの関数を用いた。データ数は相対度数化し、標本間の距離はユークリッド距離(distのeuclidean)を、またクラスター間はWard's linkage (hclustのward.D2)を指定した。
9. 匿名の査読者から、続編の文体がDickensと似ていないと判断するのに、Dickensの2作品のみとの比較では、Dickensの作品間の変異の可能性が不明なため不十分、との指摘をいただいた。文体の類似性について確定的結論を導くにはDickensの多くの作品との比較が必要だが、本論では時間的制約からEDと執筆時期に近いOMFを比較対象とした。重要な指摘であり今後の課題としたい。

10. 実際にスキャンしたのは、1874年の版である。
11. EBook#883, Release 2006/4/27, Last Updated 2016/9/25(元となった版の出版社名, 刊行年は電子テキストに記載されていない)
12. 引用符による括りで発話部を切り分けており、引用符を伴わない Free Direct/ Indirect Discourse の箇所は地の文に、また地の文に含まれる、強調のために引用符で括られた語句は発話部に切り分けている。なお発話部を機械的に「*」で検索したのでは、閉じの引用符を欠くなど括りの始端と終端が整っていない箇所があると、それ以降のテキストが正しく切り分けられない。そこで検索結果を目視で確認して、欠けた箇所や余剰な箇所に「仮の引用符; 後日識別できるよう、例えば+」を挿入するなどして、括りを整備した後に一括して切り分けた。
13. AntBNC Lemma List (antbnc_lemmas_ver_001) Laurence Anthony's Website, URL: <http://www.laurenceanthony.net/software/antconc/>
14. 本研究の対象となる3作品は、いずれも3人称語りの作品であるが、通例に従い除外した。なお、発話部についても除外した。
15. clustering height は hclust で出力された height データおよび共表形行列から得た。(ページ数の制限があるため本論には結果のみを記す)
16. hclust 分析の都度、処理対象語について相対頻度化しているため、セクション間距離への影響度の高い頻度上位語の相対頻度値は、分析語数が増えることで小さくなり、その結果クラスター高が減少する。
17. 分析データは、テキスト全体語彙の上位500語について、3つの語彙データの生起頻度を1つに集約したもの。(地の文・発話について、それぞれ少なくとも上位300語が含まれている。)
18. 注16と同じ
19. 本節では上位1,200語までの語彙数で相対度数化したうえで、1,000語までの各100語セグメントで分析しているのので、前節とは、同じ上位100語の分析結果でも、寄与率、クラスター高などの値に多少差がある。
20. ランク1-25に含まれる“s”の頻度には、*is*, *has* などの縮約によるものとs属格のサフィックスを含んでいる。そのため元来別の語彙の頻度が合計されており、今後の研究では注意して扱う必要がある。なお、この“s”の頻度データを除外して分析した場合、クラスター高などの数値はごくわずかに変化するが、クラスタリングに差異はない。
21. 本研究では語彙嗜好を際立たせるため、語彙頻度をレマでカウントしており、分析した語彙データに表記形別の頻度は含まれていない。この伝達動詞4語の頻度を同じコーパスで表記形別に再カウントしたところ、*say*, *return*, および *ask* は、EDで過去形が73%を占め、続編で逆に現在形が69%を占めている。一方、*reply* はED・続編ともに現在形が多いものの、続編では90%を占め、他の3語と同様に現在形が優勢となっている。ここから、EDと続編の語りの時制には現在形と過去形が混在しているが、それらの比がEDと続編で逆転している可能性がうかがえる。

参考文献

- B., W. H. (1874). "Review. Part Second of *The Mystery of Edwin Drood*." *The Southern Magazine*, 14 (February 1874): pp. 219-23.
- Burrows, John (1987) *Computation into Criticism – A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.
- Burrows, John (2002) "'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing*, 17(3): pp. 267-87.
- Cox, Don Richard (1998). *Charles Dickens's The Mystery of Edwin Drood : An Annotated Bibliograph*. New York: AMS Press, Inc.
- Dickens, Charles [& T. P. James] (1873). *Part Second of the Mystery of Edwin Drood. By the Spirit-Pen of Charles Dickens, Through a Medium. Embracing, Also That Part of the Work Which Was Published prior to the Termination of the Author's Earth-Life*. Brattleboro, VT.:T. P. James.
- Doyle, Arthur Conan (1927). "The Alleged Posthumous Writings of Great Authors." *The Bookman* 66, New York. [Online].
URL: <https://www.unz.org/Pub/Bookman-1927dec-00342>
- Gadd, George F. (1905). "The History of a Mystery. A Review of the Solutions to 'Edwin Drood' (Chs. 3 & 4)." *The Dickensian*, 1: pp. 270-73.
- Hill, Ralph Nading (1961). *Contrary Country: A Chronicle of Vermont*. Brattleboro, Vermont: The Stephen Greene Press.
- Hoover, David L. (2001) "Statistical Stylistics and Authorship Attribution: an Empirical Investigation." *Literary and Linguistic Computing*, 16(4): pp. 421-44.
- Hoover, David L. (2002) "Frequent Word Sequences and Statistical Stylistics." *Literary and Linguistic Computing*, 17(2): pp. 157-80.
- Hoover, David L. (2003a) "Frequent Collocations and Authorial Style." *Literary and Linguistic Computing*, 18(3): pp. 261-86.
- Hoover, David L. (2003b). "Multivariate Analysis and the Study of Style Variation." *Literary and Linguistic Computing*, 18(4): pp. 341-60.
- Hoover, David L. (2004). "Testing Burrows's Delta." *Literary and Linguistic Computing*, 19(4): pp. 453-75.
- 田畑智司 (2016) 「共著作品における Dickens の文体」堀正広・赤尾一郎 (監修), 堀正広 (編) 『英語コーパス研究シリーズ 第5巻 コーパスと英語文体』: pp. 53-71. 東京: ひつじ書房.
- Walters, John Cuming (1913) *The Complete Mystery of Edwin Drood: History, Continuations, and Solutions*. Boston: Dana ESTES & Company.
- Wolkomir, Richard (1973). "Charles Dickens' Great Mystery." *Psychic*, 4: pp. 16-17.

「論文」

英語における live a/an Adj life と lead a/an Adj life に 見られる意味的違いについて*

金澤 俊吾

Abstract

In English, *live a/an Adj life* (LIVE), classified as one of the instances of cognate object constructions, is similar to *lead a/an Adj life* (LEAD). In fact, adjectives such as *normal* and *quiet* can occur in both phrases. However, in the previous studies, the semantic properties of these phrases have not been investigated in detail. In the present study, the examples of both LIVE and LEAD are collected from the Corpus of Historical American English (COHA). To identify the date of the first occurrence of instances of each phrase, we use *the Oxford English Dictionary Online (OED Online)* and Google Books: American English, as well as COHA, classifying the instances into six semantic categories. From the results of the present research, the formation processes of each phrase are explained on the basis of ‘amalgamate’ proposed by Goldberg (1995). We claim that the formation process of LIVE should be different from that of LEAD with respect to the argument structure of each verb phrase. LIVE describes the situation where people pass their life. This phrase is characterized by an adjective, in contrast to their daily life. LEAD, on the other hand, represents the situation where people experience their life. This phrase is defined by an adjective, related to other states or their daily life.

1. はじめに

英語における動詞句 *live a/an Adj life* (以下, LIVE) から構成される同族目的語構文と, 動詞句 *lead a/an Adj life* (以下, LEAD) は, ほぼ同義として扱われている。LIVE と LEAD は, いずれも目的語に修飾句を伴い, 人生や生活の在り方を表す¹。例えば, 形容詞 *quiet* は, いずれの動詞句にも生起できる²。

- (1) a. Earl has to *live a quiet life*, ... (COHA, 1956, FIC)
 b. I tried to *lead a quiet life* ... (COHA, 1956, FIC)

実際のところ、LIVE、LEAD それぞれに関わる動詞の辞書上の定義を確認しても、違いがほとんど見られない。

- (2) i. live: to have a particular type of life, or live in a particular way
 (LDOCE⁶, s.v. live¹, v.6.)
 ii. lead: if you lead a particular kind of life, that is what your life is like.
 (LDOCE⁶, s.v. lead¹, v.10.)

しかし、(3) に示すように、形容詞 beautiful は、LIVE には生起できるが、LEAD に生起できる例は見つけられない。

- (3) We want to *live a beautiful life* ... (COHA, 1941, FIC)

LIVE、LEAD が類似した表現でありながら、なぜ各動詞句に生起する形容詞の分布に違いが見られるのであろうか。本稿では、The Corpus of Historical American English (COHA) に見られる LIVE、LEAD に生起する形容詞の分布の意味的推移と、各動詞句の形成過程を考察することで、この問いに対する説明を試みる。

2. 先行研究における LIVE、LEAD について

LIVE に関する先行研究において、Höche (2009) は、live a happy life の例を挙げ、同族目的語と様態の副詞との意味的振る舞いの違いを指摘する。live a happy life は、「一定の期間、幸せな生活を過ごす」と解釈されるのに対し、live happily は、「幸せに生活する」と解釈される。また、小西 (2001: p. 939) は、形容詞を伴わない動詞句 live a life を挙げ、受動化の可能性を指摘する³。

LEAD に関して、Wood (1964) は、二重目的語構文に lead が生起し、直接目的語として a life を取る場合、修飾する形容詞は「不快もしくは望まない生活」を表す内容に限定されることを指摘する (He led his wife a miserable life. (Wood, 1964: p. 161))。さらに、LIVE と LEAD の意味的特徴に関して、Wood (1964)

と小西(1980)によると, LIVE は単に日々の生活を過ごす状況を表すのに対し, LEAD は意識的かつ積極的に生活を営む状況を表す点に違いが見られる。

いずれの研究も, LIVE と LEAD の形成過程が詳細に検討されておらず, 両動詞句の意味的特徴が十分に考察されているとは言えない状況にある。

3. COHA に見られる LIVE と LEAD について

3.1 LIVE, LEAD の用例の調査方法について

本研究は, いわゆる探索型調査の立場で進めていく。LIVE, LEAD の十分な量の用例を調査し検証するため, 4億語が収録される COHA を使用する。はじめに, (4) に示す検索式を用いてレマ検索を行う。その検索結果から得られる不定冠詞 a, an と, 名詞 life との間に形容詞 1 語を含む, 100 万語あたり 0.01 以上, 使用頻度 3 以上の LIVE と LEAD の用例を考察の対象とする⁴。

(4) [live][a]* life [lead][a]* life

その後, 各動詞句に生起する形容詞の初出年を調査し, LIVE, LEAD に生起する形容詞の分布の意味的推移を検証する。

当該表現の初出年の特定をより精緻化するため, COHA に加え, 英語の歴史に基づいて編纂されている *The Oxford English Dictionary Online (OED Online)* と, COHA で調査する LIVE, LEAD が, 小説に多く見られる動詞句であり, アメリカ英語であることから, アメリカ英語でデータ・サイズが大きい Google Books: American English を用いる。OED Online と Google Books: American English を使用する理由は 2 つある。第 1 に, COHA には 1810 年から 2009 年までのデータが収録されているので, 1810 年以前の LIVE と LEAD に生起する形容詞の初出年を確認するためには, これらを用いて補完しなければならないからである。第 2 に, COHA において, 一方の動詞句のみに生起する形容詞であっても, 他のコーパスでは他方の動詞句に生起する可能性が想定されるためである。

本研究において, LIVE, LEAD の初出年の調査に基づき, 形容詞の意味的推移を考察することで, 両動詞句が, どのような通時的形成過程を経て, 現在ほぼ同義の動詞句であるかを解明できるという利点が挙げられる。ただし, 今回, 考察の対象が, COHA に収録された用例に限定されるので, 当該動詞句の形

成過程の解明が部分的であるという点において、限界が見られる。

3.2 LIVE, LEAD の使用頻度について

先述の検索式を用いて COHA を調査した結果、LIVE と LEAD のタイプ数と、使用頻度は、表 1 に示す結果が得られた。

表 1 COHA における LIVE, LEAD のタイプ数と使用頻度

	タイプ数	使用頻度
LIVE	31	256
LEAD	39	343

また、LIVE に生起する 256 例のうち、各形容詞の使用頻度と、当該動詞句に占める割合は、表 2 としてまとめられる。

表 2 COHA における LIVE の形容詞の使用頻度と当該動詞句に占める割合

	形容詞	使用頻度	割合(%)
1	good	30	11.7
2	normal	28	10.9
3	new	16	6.3
4	double	14	5.5
5	full	14	5.5
6	better	13	5.1
7	quiet	13	5.1
8	lonely	12	4.7
9	simple	11	4.3
10	healthy	9	3.5
11	perfect	9	3.5

	形容詞	使用頻度	割合(%)
12	hard	8	3.1
13	solitary	8	3.1
14	real	7	2.7
15	wild	6	2.3
16	happy	6	2.3
17	single	5	2.0
18	secluded	5	2.0
19	different	4	1.6
20	useful	4	1.6
21	wandering	4	1.6
22	complicated	3	1.2

	形容詞	使用頻度	割合(%)
23	miserable	3	1.2
24	dangerous	3	1.2
25	bad	3	1.2
26	beautiful	3	1.2
27	comfortable	3	1.2
28	peaceful	3	1.2
29	pure	3	1.2
30	free	3	1.2
31	common	3	1.2

LIVE に生起する上位 3 位まで見ると、good, normal, new が生起し易いことが明らかとなる。さらに、LEAD に生起する 343 例は、各形容詞の使用頻度と、当該動詞句に占める割合は、表 3 にまとめられる。

表3 COHAにおける LEAD の形容詞の使用頻度と当該動詞句に占める割合

	形容詞	使用頻度	割合(%)
1	double	43	12.5
2	normal	33	9.6
3	different	31	9.0
4	good	20	5.8
5	better	19	5.5
6	quiet	19	5.5
7	new	17	5.0
8	happy	12	3.5
9	lonely	12	3.5
10	sheltered	12	3.5
11	solitary	8	2.3
12	roving	8	2.3
13	wandering	8	2.3

	形容詞	使用頻度	割合(%)
14	pleasant	7	2.0
15	dual	5	1.5
16	comfortable	5	1.5
17	single	5	1.5
18	hard	4	1.2
19	miserable	4	1.2
20	precarious	4	1.2
21	wild	4	1.2
22	rough	4	1.2
23	merry	4	1.2
24	full	4	1.2
25	busy	4	1.2
26	perfect	4	1.2

	形容詞	使用頻度	割合(%)
27	simple	4	1.2
28	private	4	1.2
29	vagabond	4	1.2
30	secret	4	1.2
31	bad	3	0.9
32	vile	3	0.9
33	cat-and-dog	3	0.9
34	useful	3	0.9
35	pure	3	0.9
36	regular	3	0.9
37	sedentary	3	0.9
38	tranquil	3	0.9
39	nomadic	3	0.9

LEAD に生起する上位 10 位までの形容詞のうち、6つの形容詞が、LIVE の分布と重なることが分かる (double, normal, good, better, quiet, new)。また、表 4 は、LIVE, LEAD いずれにも生起する形容詞の使用頻度を示している。

表4 COHAにおける LIVE と LEAD いずれにも生起する形容詞の使用頻度

	形容詞	LIVE	LEAD
1	good	30	20
2	normal	28	33
3	new	16	17
4	double	14	43
5	full	14	4
6	better	13	19
7	quiet	13	19
8	lonely	12	12
9	simple	11	4
10	perfect	9	4
11	hard	8	4

	形容詞	LIVE	LEAD
12	solitary	8	8
13	wild	6	4
14	happy	6	12
15	single	5	5
16	different	4	31
17	useful	4	3
18	wandering	4	8
19	miserable	3	4
20	bad	3	3
21	comfortable	3	5
22	pure	3	3

2つの動詞句いずれにも生起する形容詞のタイプ数は22あり、LIVEのタイプ数の71.0%を、LEADのタイプ数の56.4%を占める。この重なりの高さは、両動詞句に生起する形容詞の分布が、互いに類似していることを強く示唆している。

3.3 LIVE, LEAD に生起する形容詞の分布の推移について

3.2節の調査によって得られた各動詞句に生起する、全ての形容詞を調査対象とし、当該の形容詞の初出年を調査し、各動詞句に生起する、形容詞の意味的分布の推移を考察する。各動詞句に生起する形容詞を、初出年の早い順に並べると、各動詞句に類似した概念的特徴を持つ形容詞の分布が見られる。

本稿では、a life を修飾する形容詞は、Langacker (1987) の言う、名詞によって表される実体（ここでは a life）を関係づけ、状態性を表すことで、その実体の特徴をプロファイルする機能を担うと考える。また、Paradis (2001) による、形容詞の語彙的意味には、動詞と同様、境界性 (boundedness) が内在していると考え、LIVE, LEAD の意味的特徴を解明する上で重要な役割を担うと考える。さらに、Dixon (1982) による、語レベルの形容詞の意味タイプの分類を参考にする。最終的に、調査対象全ての形容詞が、いずれかのカテゴリーに分類されるよう、(5) に挙げる6つのカテゴリーを設定する⁵。

(5) DIFFERENCE, ONE, WANDER, MANNER, GOOD, BAD

はじめに、初出年が早い形容詞を含むカテゴリーの順に従い6つのカテゴリーを紹介する。LIVE と LEAD とともに最も古くに初出が確認できる形容詞は new であり、1584年に初出する。この形容詞を含むカテゴリーを DIFFERENCE とし、日常生活とは異なる生活を表す4つの形容詞が分類される⁶。

表5 DIFFERENCE に分類される、LIVE, LEAD に生起する形容詞の初出年

	形容詞	LIVE	LEAD		形容詞	LIVE	LEAD
1	new	<u>1584(G)</u>	<u>1584(G)</u>	3	double	<u>1851(C)</u>	1854(G)
2	different	1829(C)	<u>1814(G)</u>	4	dual	<u>1873(G)</u>	1875(G)

このカテゴリーに分類される形容詞は、LIVE, LEAD いずれの動詞句にもほぼ同時期に初出する。new 以外にも、1800年代に入ると different, double, dual が、

いずれの動詞句にもほぼ同時期に初出する。

DIFFERENCE と並びほぼ同時期に、初出が確認できる形容詞が属するカテゴリーとして、ONE が挙げられる。このカテゴリーには、3つの形容詞 (single, solitary, lonely) が分類され、いずれも1人で生活する様子を表す。

表6 ONE に分類される、LIVE, LEAD に生起する形容詞の初出年

	形容詞	LIVE	LEAD
1	single	<u>1584(G)</u>	1608(G)
2	solitary	1810(G)	<u>1697(O)</u>
3	lonely	1851(G)	<u>1849(G)</u>

single は、new と同年(1584年)にLIVEに初出後、LEADに初出する(1608年)。また、solitary は、1697年にLEADに初出後、1810年にLIVEに生起する。さらに、lonely は、LIVE と LEAD いずれにも1850年前後に初出する。

次に、LIVE, LEAD への初出に時間的先行関係が顕著に見られる、4つのカテゴリーを紹介する。WANDER がこれに該当し、世俗を離れての生活状況を表す8つの形容詞がこのカテゴリーに分類される。

表7 WANDER に分類される、LIVE, LEAD に生起する形容詞の初出年

	形容詞	LIVE	LEAD
1	private	1755(G)	<u>1636(O)</u>
2	wandering	1808(G)	<u>1785(G)</u>
3	roving	1854(G)	<u>1801(G)</u>
4	vagabond	1852(G)	<u>1821(G)</u>

	形容詞	LIVE	LEAD
5	nomadic	<u>1830(G)</u>	1834(G)
6	secluded	<u>1832(G)</u>	1833(G)
7	secret	<u>1865(G)</u>	1913(G)
8	sheltered	<u>1882(G)</u>	1901(G)

ここで特筆すべき点は、1820年代頃まではLEADに初出後、LIVEに初出する形容詞 (private, wandering, roving, vagabond) と、1830年以降、LIVEに初出後、LEADに初出する形容詞 (nomadic, secluded, secret, sheltered) が見られることである。

WANDER に次いで、初出の早い形容詞が属するカテゴリーは、MANNER であり、生活の様子や様態を表す15の形容詞がこのカテゴリーに分類される。

表8 MANNER に分類される, LIVE, LEAD に生起する形容詞の初出年

	形容詞	LIVE	LEAD
1	regular	1803(G)	<u>1777(O)</u>
2	sedentary	1851(G)	<u>1802(G)</u>
3	perfect	1845(G)	<u>1811(G)</u>
4	full	1864(G)	<u>1811(G)</u>
5	peaceful	1833(G)	<u>1812(G)</u>
6	quiet	<u>1811(G)</u>	1821(G)
7	tranquil	<u>1827(G)</u>	1833(G)
8	busy	1871(G)	<u>1835(G)</u>

	形容詞	LIVE	LEAD
9	pure	<u>1835(C)</u>	1845(G)
10	free	<u>1841(G)</u>	1851(G)
11	common	<u>1844(C, G)</u>	1865(G)
12	simple	<u>1851(G)</u>	<u>1851(G)</u>
13	healthy	<u>1861(G)</u>	1871(G)
14	real	<u>1863(C)</u>	1895(G)
15	normal	<u>1871(G)</u>	1894(G)

18世紀後半から19世紀前半にかけて、LEADに初出後、LIVEに初出する形容詞が見られ、生活の全容を表す形容詞がそれに該当する (regular, sedentary, perfect, full)。一方、LIVEに初出後、LEADに初出する形容詞があり、1810年以降、生活を平穩に過ごす状況を表す形容詞 (quiet, tranquil, pure) が、1840年以降、通常的生活状態を表す形容詞 (common, simple, healthy, real, normal) が、それぞれ初出する。

次に、MANNER とほぼ同時期に初出が見られるカテゴリーとして、価値判断を表す GOOD が挙げられる。このカテゴリーには8つの形容詞が属し、話者や主語の主観的判断に基づき、生活の豊かさを表している。

表9 GOOD に分類される, LIVE, LEAD に生起する形容詞の初出年

	形容詞	LIVE	LEAD
1	good	1832(G)	<u>1792(G)</u>
2	better	1821(G)	<u>1801(G)</u>
3	happy	1833(G)	<u>1804(G)</u>
4	merry	<u>1802(G)</u>	1834(G)

	形容詞	LIVE	LEAD
5	pleasant	<u>1810(G)</u>	1833(G)
6	comfortable	<u>1816(G)</u>	1831(G)
7	useful	1847(C)	<u>1835(G)</u>
8	beautiful	<u>1863(G)</u>	—

LEADに初出後、LIVEに初出する形容詞 (good, better, happy, useful) と、LIVEに初出後、LEADに初出する形容詞 (merry, pleasant, comfortable, beautiful) が見られる。また、初出が1830年代に集中的に見られる形容詞があり、LIVEではgood (1832年) が、LEADではmerry (1834年) とuseful (1835年) が該当する。

最も遅くに初出が確認されるカテゴリーは、価値判断を表す形容詞のうち、生活が芳しくない状況を表す BAD であり、10 の形容詞が分類される。

表10 BAD に分類される、LIVE, LEAD に生起する形容詞の初出年

	形容詞	LIVE	LEAD
1	bad	1833(G)	<u>1811(G)</u>
2	miserable	1833(G)	<u>1822(G)</u>
3	precarious	1844(G)	<u>1822(G)</u>
4	wild	<u>1831(G)</u>	1835(G)
5	hard	<u>1851(G)</u>	<u>1851(G)</u>

	形容詞	LIVE	LEAD
6	rough	<u>1853(G)</u>	1854(G)
7	vile	<u>1854(G)</u>	<u>1854(G)</u>
8	cat-and-dog	—	<u>1874(C)</u>
9	dangerous	<u>1926(G)</u>	<u>1926(G)</u>
10	complicated	<u>1927(G)</u>	—

LEAD が LIVE に先行して初出する形容詞 (bad, miserable, precarious) が見られ、その後、1830 年以降、LIVE に初出後、LEAD に初出する形容詞が見られる (wild)。また、2 つの動詞句にほぼ同時期に初出する形容詞 (hard, rough, vile, dangerous) や、LEAD にのみ生起する形容詞 (cat-and-dog)、LIVE にのみ生起する形容詞 (complicated) が見られる。

この初出年に基づく調査結果から、以下の3点が明らかとなる。第1に、LIVE に初出後、LEAD に生起する形容詞と、LEAD に初出後、LIVE に生起する形容詞が体系的に存在することを確認できる。この形容詞の分布の広がりから、LIVE, LEAD は生起する形容詞の分布に相互に影響を与えながら、各動詞句が形成され、生産性が高められていることが明らかとなる。

第2に、(6) に示すように、LIVE, LEAD に生起する形容詞の分布は、生活を客観的に表す形容詞から、主観的に表す形容詞に至るまで、3段階に分かれて意味的推移が見られることが明らかとなる。

- (6) i. 第1段階：他の生活との対比、他者から離れた生活状況を端的に表す。
 ii. 第2段階：世俗から離れた人生、生活の過程を表す。
 iii. 第3段階：話者、主語の価値判断に基づく生活状況を表す。

形容詞の初出が最も早い第1段階では、DIFFERENCE と ONE にそれぞれ分類される形容詞が、各動詞句に生起し、対比される生活状況が明示されることによって、生活状況が客観的に表される。第2段階では、MANNER に分類される形容詞が生起することで、対比される生活が明示されることなく、世俗から

離れた人生や、生活の過程が表される。その後、第3段階において、GOODとBADに分類される形容詞によって、話者や主語の価値判断に基づく、主観的な生活状況が表される。Traugott (1995) は、文法化の過程において、当初、客観的な状況を表す語が生起していたものが、主観的な状況を表す語の生起を許容できるようになる、主観化 (subjectification) が見られることを指摘する。このLIVE, LEADに生起する形容詞の意味的推移には、文法化は見られないが、客観的な状況を表す形容詞から、主観的な状況を表す形容詞へと分布が推移し拡大する点において、この「主観化」が反映されていると言える。

第3に、COHAから得られた形容詞の初出年と使用頻度との間に相関性が見られるカテゴリーと、ほとんど見られないカテゴリーがあることが明らかとなる。DIFFERNCEとONEには、この相関性が見られ、当該カテゴリーに分類される形容詞は、2つの動詞句にほぼ同時期に生起し、いずれも頻度が高い。一方、WANDERと、MANNER, GOOD, BADの4カテゴリーに分類される形容詞には、この相関性がほとんど見られない。初出年の時間的先行関係に関わらず、時間の経過とともに、LIVEに生起し易い形容詞のカテゴリー (GOOD) と、LEADに生起し易い形容詞のカテゴリー (WANDERとBAD) が存在する。また、MANNERに分類される形容詞は、LIVEには生活全般を表す形容詞 (normal, full) や、平穏な生活状況を表す形容詞 (quiet, healthy, peaceful, common) が生起し易い。一方、LEADには、具体的な生活の動きの様子を表す形容詞 (busy, sedentary) が生起し易い。

なぜ、2つの動詞句に生起する形容詞の分布に、このような違いが見られるのであろうか。これは、LIVE, LEADの形成過程の違いに起因すると考えられる。4節では当該動詞句の形成過程を考察し、5節において、各動詞句に生起する形容詞の意味的分布の違いに対する説明を試みる。

4. LIVE, LEADの形成過程について

本稿では、LIVE, LEADは、いずれの場合においても2つの動詞の意味が融合 (amalgamate) されることで形成される動詞句であり、この融合の違いが、各動詞句に生起する形容詞の分布の違いに反映されると主張する。

LIVEに生起する動詞 live は、本来、自動詞であり、名詞句 a/an Adj life が目的語位置に生起することで他動詞のように振る舞う。この文法的振る舞いを考える上で、Goldberg (1995) による One's Way 構文の形成過程に関する考察

が有効である。

(7) He made his way into the room. (Goldberg, 1995: p. 207)

- i. He made a path.
- ii. He moved into the room.

One's Way 構文は、2つの構文、すなわち、他動詞 make が生起する経路作成の構文((7i))と、自動詞 move が生起して前置詞句によって移動を表す構文((7ii))が融合されることで形成される。最終的に、(7)は、「彼」が道を切り開いて進み、部屋へたどり着くと解釈される。この融合の結果、One's Way 構文には他動詞 make のみならず、wend, crack などの自動詞も生起できる。

本稿では、この Goldberg (1995) による融合の知見を援用し、LIVE は、自動詞構文 live 'live in a particular way' と、他動詞構文 pass 'pass time or pass your life in a particular way' が融合されることで形成されることを提案する。

(8) He lived a happy life.

- i. He passed time in a happy way.
- ii. He lived in a happy way.

動詞 pass の目的語位置に生起する、過ごす時間を表す名詞句 time が a life として具現化される。また、動詞 pass, 動詞 live の語彙的意味に指定される様態 ('in a happy way') は、形容詞 happy によって具現化され、a life を修飾する。

その結果、LIVE は、「(通常の生活と比較、対比されることで) ある特定の状態で生活を過ごす」という意味を表す。LIVE が、特定の状態で時間を過ごすとして解釈されることは、Höche (2009) による「同族目的語は、動詞によって表される動作が一定期間に及ぶ状況を記号化する」という指摘からも支持される。

次に、LEAD の形成過程について考察する。本稿では、LEAD は、動詞 lead に指定される、2つの語彙的意味の融合により形成されると提案する。動詞 lead は、lead₁ から lead₄ へ意味拡張が見られる (lead の各定義は、LDOCE⁶ による)。

(9) lead₁: go in front → lead₂: be in charge → lead₃: influence someone to make them do something that is wrong → lead₄: be more successful than other people etc. in a particular activity

lead₁ は、「先頭を切って前に進む」という意味を表す。その先頭を切る対象が抽象的な実体に変化することで、動詞 lead₂ 「ある事柄に対して、責任がある、管理する」の意味を表す。また、lead₃ は、二重目的語構文を用いて「誰かに対して悪い影響を与えるよう働きかける」状況を表す。その際、主語には「使役主」(causer) の意味役割を持つ項が、直接目的語には悪い状況を表す比喩的な「経路」(path) の意味役割を持つ項が、それぞれ生起する。また、間接目的語には、その影響を受ける「主題」の意味役割を持つ項が生起する。さらに、lead₄ に意味拡張が進むと、他の状態との比較、対比が含意され、ある特定分野で成功を収める状態を表す。

本稿では、LEAD は、lead₃ の意味の一部と lead₄ の意味が融合することで形成されると提案する。具体的には lead₃ における「使役主」が捨象され、「主題」が結果状態に至るまでの過程を表す部分と、lead₄ に見られる、他の状態との比較、対比によって生じる、目立つ状況を表す意味が融合されることで形成されると考える。最終的に、LEAD は「ある特定の状態、状況との比較、対比により、形容詞によって表される状態で生活を過ごす」という意味を表す。

5. LIVE, LEAD に生起する形容詞の意味的振る舞い

4 節で見た、動詞句 LIVE と LEAD の形成過程の違いに基づき、LIVE, LEAD にそれぞれ生起し易い形容詞のカテゴリーが見られる理由を説明する。

5.1 LIVE に生起する形容詞の意味的振る舞い

LIVE には、GOOD と、MANNER の中でも生活全般を表す形容詞 (normal, full) や、平穏な生活状況を表す形容詞 (quiet, healthy, peaceful, common) が生起し易いという意味的特徴が見られた。LIVE は、「(通常の生活と対比、比較することで) ある特定の状態で生活を過ごす」という意味を表す。また、談話的特徴として、生活状況や生活の様子が包括的に提示された上で、後続する文や句によって、当該の生活が詳述される。

この状況下で、LIVE に生起する形容詞は、通常の生活と対比されることで、包括的に提示される生活を具現化する。この意味的振る舞いが、LIVE の状態性の高い意味（「時間を過ごす」）との間で整合性が図られ、その結果、GOOD や MANNER に属する一部の形容詞が、LIVE に生起し易いと説明される。

LIVE に見られるこの意味的特徴に基づき、意味的推移の第 1 段階に位置づ

けられ、DIFFERENCE に分類される、形容詞 new の具体例を見てみよう。

- (10) She wants to *live a new life*, is working hard, and is trying to break away from smoking. (COHA, 1909, FIC)

(10) における a new life は、通常の生活との対比により、異なる生活状況を包括的に表し、「彼女」が新たな生活を送ることを望んでいると解釈される。その「新たな生活」が、後続する「一生懸命になって働く」事象と、「たばこを断とうとする」事象によって、それぞれ詳述される。

次に、意味的推移の第2段階に位置づけられ、MANNER に分類される形容詞 quiet が LIVE に生起する例を説明する。具体例として (11) を見る。

- (11) They are gloomy and melancholic animals; they *live a quiet life*, climbing trees, and eating fruit and insects. (COHA, 1897, NF)

quiet によって、「彼ら」が過ごす平穏な生活が包括的に表される。その「平穏な生活」が、後続する動詞句により、木登りと、果物や昆虫を食べることで営まれていると詳述される。

また、意味的推移の第3段階に位置し、話者や主語名詞による、生活に対する価値判断を表す GOOD についても同様に説明される。(12) の a good life を例に考えてみよう。

- (12) ... Dazzle *lived a good life* in the woods. He ate plenty of fresh fruit and vegetables, learned to take one day at a time ... (COHA, 1999, FIC)

(12) では、Dazzle が森の中で豊かな生活をしている様子が、a good life によって包括的に描写されている。その上で、後続する2つの動詞句によって、豊富な果物を食べて、その果物を1日で同時に取れるようになったと詳述される。

5.2 LEAD に生起する形容詞の意味的振る舞い

LEAD には、WANDER, BAD にそれぞれ分類される形容詞や、MANNER の一部の形容詞 (busy, sedentary) が生起し易いことを指摘した。LEAD は、「ある特定の状態、状況との対比、比較により、形容詞によって表される生活を過

ごす」という意味を表し、「不快もしくは望まない生活」を過ごす状況を表す傾向が強い。さらに、談話の特徴として、対比される事象が、LEADによって表される事象に先行して提示され、因果関係もしくは対比関係が構築される。この点で、LIVE に比べ、LEAD は、より動的に解釈される。

また、LEAD に生起する形容詞は、先行する他の事象との対比によって、a life によって表される生活時間の過程を具現化する。形容詞のこの意味的振る舞いと、LEAD の意味との間で整合性が図られることにより、WANDER や BAD, MANNER の一部に属する形容詞が生起し易いと説明される。

LIVE と同様、LEAD にも、(13) のように、意味的推移の第1段階に位置づけられる、DIFFERENCE に分類される形容詞 *new* が生起できる。

- (13) I will go home and humble myself, and reflect upon my past follies, and try to repent, and tomorrow I will *lead a new life*. (COHA, 1845, FIC)

ただし、LIVE とは異なり、(13) は、先行する事象と因果関係が構築されることによって *lead a new life* が表されている。家に帰り、謙虚になって、過去の愚行を熟考し、後悔するという一連の事象が列挙される。その上で、*lead a new life* によって、これまでの生活を変えることで、新たな生活を送る状況が表される。

次に、意味的推移の第2段階に位置づけられ、WANDER に分類される *wandering* を見る。(14) を例に考えてみよう。

- (14) The inland savages, on the other hand, *led a wandering life*, [...] only coming down occasionally to rob the plantations of the coast tribes ... (COHA, 1863, NF)

この文の先行文脈では、ポルトガルからの移民を好意的に受け入れた、性格の穏やかな人々の生活の様子が表されている。この生活状況と対比されることで、*inland savage* が放浪生活をしている様子が表される。また、MANNER についても、WANDER と同様、他の生活状況と対比されることで、過ごす生活の様子が具現化される。(15) の形容詞 *sedentary* を例に考えてみる。

- (15) Small, heavy-set pilots are more resistant to blackout than tall, slender men, and those *leading a sedentary life* have more resistance than men in athletic training. (COHA, 1941, MAG)

運動して鍛えているパイロットとの対比により、運動することなく、じっと生活している人の方が、一時的に意識を失うことに対して耐性がある状況が描写される。

最後に、意味的推移の第3段階に位置づけられる BAD の具体例を見る。(16) の bad は、他の状況との対比によって、「芳しくない生活」を描写する。

- (16) A sick lady came there and the Sisters said she used to be beautiful [...] she was going to die alone and unloved because she'd *led a bad life*. (COHA, 1949, FIC)

「彼女」が、かつて舞台の花形で活躍していた状況との対比によって、晩年、不遇な人生を送ってきた様子が描写される。さらに、不遇な人生の結果、愛されることなく、1人で亡くなりつつある状況が描写されている。

6. おわりに

英語の動詞句 LIVE, LEAD の意味的特徴について、当該動詞句に生起する形容詞の初出年を調査し、当該の形容詞のカテゴリーの分布の推移と、各動詞句の形成過程の違いに基づき考察した。その結果、LIVE, LEAD は、各動詞句に生起する形容詞の分布に関して、互いに影響を与えながら拡大することにより形成されていることを明らかにした。また、各動詞句に生起する形容詞のカテゴリーには、生活状態を、客観的に表す形容詞から、主観的に表す形容詞に至るまで、3段階に分かれて推移、拡大し、「主観化」が見られることを明らかにした。さらに、各動詞句に生起する形容詞の初出年の時間的先行関係にかかわらず、LIVE に生起し易い形容詞と、LEAD に生起し易い形容詞が存在することを指摘し、これは、動詞句の LIVE, LEAD の形成過程の違いに起因していることを明らかにした。とりわけ、LIVE は、通常的生活、これまでの生活との対比の中で、形容詞によって生活が包括的に特徴づけられ、後続する句や文によって、その具体的な生活が詳述される意味的特徴を有することを示した。一方、LEAD は、他の事象との対比により、因果関係ないし対比関係が構築されることで、形容詞によって表される生活を経験する状況を表すという意味的特徴を有することを示した。

注

* 本稿の内容の一部は、2017年8月28日、東北大学大学院情報科学研究科「言語変化・変異研究ユニット」主催第4回ワークショップ「コーパス・多人数質問調査からわかる言語変化・変異と現代言語理論」において口頭発表した、「英語における同族目的語構文の形成過程とその推移について—live a/an Adj life と lead a/an Adj life を中心に一」に基づいている。また、匿名の査読者より、有益かつ的確な質問、助言を賜った。ここに謝意を表したい。本稿における誤り等に関する責任は、すべて筆者にある。なお、本研究の成果の一部は、JSPS 科研費（課題番号：16K02774）の助成を受けてなされている。

1. 前置詞 of を用いて a life を修飾するボタン (lead a lie of peace) も見られるが、本稿では、形容詞が名詞句 a life を限定修飾する事例に限定して議論する。
2. 以下、用例中の斜字体は筆者による。
3. 形容詞を伴うことなく、動詞句 smile a smile が容認されることが、Schibsbye (1979) と大室 (1991) によって指摘されている。
4. 接尾辞 -er を伴う、形容詞の比較級も検索結果も各形容詞の使用頻度を含める。なお、better は、他の語に比べて頻度が高いため、1つの語彙項目として扱う。
5. 時間を表す形容詞 (long, whole) や、道徳性を表す形容詞 (blameless, charmed, religious など) も見られるが、今回は分析の対象外とし、今後の研究課題とする。
6. 用例中、カッコ内の数値は初出年を表す。また、初出年と併記される C (COHA), G (Google Books: American English), O (OED Online) は、いずれも用例の出典を表す。下線は、一方の表現が、他方よりも時間的に先行して初出することを示す。

参考文献

- Dixon, R. M. W. (1982) *Where Have All the Adjectives Gone?* Berlin: Mouton.
- Goldberg, A. E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
- Höche, S. (2009) *Cognate Object Constructions in English: A Cognitive-Linguistic Approach*. Tübingen: Gunter Narr Verlag.
- 小西友七 (1980) 『英語基本動詞辞典』研究社.
- 小西友七 (2001) 『英語基本名詞辞典』研究社.
- Langacker, R. W. (1987) *Foundations of Cognitive Grammar: Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- 大室剛志 (1991) 「同族‘目的語’構文の特異性 (3)」『英語教育』1991年1月号: 68-72.
- Paradis, C. (2001) “Adjectives and Boundedness.” *Cognitive Linguistics* 12, 1: 47-64.
- Schibsbye, K. (1979) *A Modern English Grammar Second Edition*. Oxford: Oxford University Press.
- Traugott, E. C. (1995) “Subjectification in Grammaticalization.” In Stein, D. and R. D. Janda (eds.), *Subjectivity and Subjectivization*. Oxford: Blackwell, pp. 624-647.

Wood, F. T. (1964) *English Verbal Idioms*. London: Macmillan.

辞書・コーパス

Longman Dictionary of Contemporary English Sixth Edition (LDOCE⁶) (2014) Harlow: Pearson Education.

The Oxford English Dictionary Online (OED Online): <http://www.oed.com/>

The Corpus of Historical American English (COHA): <https://www.english-corpora.org/coha/>

Google Books: American English: <https://googlebooks.byu.edu/>

(高知県立大学 E-mail: kanazawa@cc.u-kochi.ac.jp)

「論文」

The Time Span Required for Syntactic Changes in the History of English

Satoru TSUKAMOTO

Abstract

This study attempts to explore the duration of time needed for syntactic changes over time. Various changes have taken place in the history of English, such as in the choice of perfect auxiliary of mutative verbs or nominal inflection. This paper examines nine linguistic changes in English, which range from verb systems to genitive case, and adverb or adjective components, by use of parsed corpora: PPCME2, PPCEME, PPCMBE2, and YCOE. The result indicates that five of the nine linguistic changes continued for 200 or 300 years and remained stable. On the other hand, the changes which became obsolete only continued for about 100 or so years. This difference shows that a linguistic change requires a duration of time of 200 to 300 years, and that the amount of time required for changes is constant.

1. Introduction

The English language has changed in various aspects over time. Statements regarding linguistic changes have been addressed in Jespersen (1909–1949), Fries (1940), Mossé (1938), Ellegård (1953), Visser (1963–1973), and Denison (1993), to mention a few. Denison (1993) describes historical changes in English, mainly focusing on verbal constructions. He does not, however, demonstrate the frequency of the developments over time. It is not clear how long these developments continued in the history of English. Mossé (1938) illustrates the occurrences of what is called the periphrastic form by several lemma types in Middle English texts. Jespersen compares the occurrences of the expanded tense (i.e. *be -ing* construction) between numbers in the 16th-century and 20th-century Bibles, showing that the latter version includes about four times more frequent use of the construction (Jespersen, 1909: vol. IV. 177). We can recognize the

growth of the construction but it is hard to pinpoint the beginning and end of the change. Furthermore, those studies deal with a particular grammatical item in isolation, and do not provide comparable perspectives on linguistic changes.

This study aims to examine the duration of nine syntactic changes which took place over time by retrieving data from the same text sources (i.e. the same corpora), and to show that, by presenting these results in juxtaposition, linguistic developments show analogous increases over time.

2. Corpora and Method

In this study the following corpora are employed. These corpora cover almost all the periods in the history of English. They are compiled under the same principle in grammatical tagging, which makes it possible to compare linguistic occurrences between different periods. The method used here is to retrieve the specified construction marked in the parsed corpora, and to examine the rate of changes over time.

The Penn-Helsinki Parsed Corpus of Middle English (PPCME2)

The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)

The Penn Parsed Corpus of Modern British English (PPCMBE2)

The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)

In these corpora most of the texts are classified into subperiods of 70 years. However, in this study each text is grouped into 25-year periods to clearly exemplify the chronological development of the linguistic items.¹ This segmentation by 25 years enables one to detect a short-term fluctuation, i.e. an increase or decrease in a short time period, which would not be observable in a longer time span. By utilizing the corpora which include the syntactic annotations, this study focuses on the syntactic changes over the history of English. The changes to be dealt with here include:

- (1) Personalization of impersonal verb *like*
- (2) The choice of perfect auxiliary verbs with mutative verbs
- (3) The development of the progressive form
- (4) The emergence of the gerund (*-ing*) complement of verbs
- (5) Genitive forms replaced by 'of' periphrasis

- (6) The use of *thereof* against *of it*
- (7) The obligatory use of the determiner
- (8) The decline of double determiners
- (9) Double comparatives and superlatives

The linguistic items under consideration are not selected upon a particular principle. Rather, this selection is arbitrarily made in order to capture a general tendency shown in a grammatical change over time.

3. Results and discussions

3.1 Personalization of impersonal verb *like*

The impersonal verb *like* takes the argument of “experiencer” in the objective case or object position as in (1).

- (1) every knyght toke the way that *hym* lyked beste. (CMMALORY,638.3880)
No, and it lyke *your Majestie*, answered he (PERROTT-E2-P1,47.14)

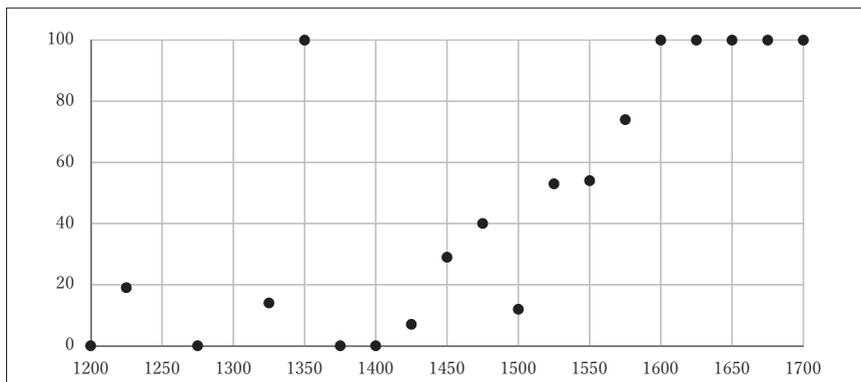
The argument can appear in the nominative case or subject position as in (2).

- (2) *The kynge* lyked and loved this lady wel (CMMALORY,2.12)
I like your wordes well (GIFFORD-E2-H,B4R.237)

In this study the impersonal use of the verb *like* is defined as the instance which takes the experiencer as object, whereas the personal use is defined as the instance which takes it as subject. This development is shown in Table 1 below.

Table 1: Personalization of impersonal verb *like*

Date	Personal Use	Impersonal Use	Personalized Use (%)	Total
1200	0	4	0	4
1225	7	29	19	36
1250	-	-	-	-
1275	0	0	0	0
1300	-	-	-	-
1325	1	6	14	7
1350	1	0	100	1
1375	0	20	0	20
1400	0	12	0	12
1425	2	25	7	27
1450	16	39	29	55
1475	2	3	40	5
1500	5	37	12	42
1525	8	9	47	17
1550	21	13	54	34
1575	15	5	74	20
1600	5	0	100	5
1625	17	0	100	17
1650	11	0	100	11
1675	21	0	100	21
1700	10	0	100	10

Figure 1: Personalization rate of impersonal verb *like*

The occurrence in personal use began to increase in about 1400 and reached the maximum in 1600, when the impersonal use of the verb disappeared. The personalization process was completed after a span of 200 years.

3.2 The choice of perfect auxiliary verbs with mutative verbs

In earlier English, the perfect form of the mutative verbs was construed with the verb *be* as an auxiliary verb, as in (3).

- (3) His mouth *is* now *become* very sore (BARDSLEY-1807-1,40.418)
 Ye *are come* unto Mount Zion, and unto the city of the living God
 (TALBOT-1901-1,89.7)
 how . . . thou *art fallen* into sickness (BOETHJA-1897-2,27.36)
 all the life *is gone* out of them (THRING-1883-2,183.332)

The choice of the auxiliary changed over time, resulting in *have*-perfect in Present-Day English. As Rydén and Brorström (1987), Rydén (1991) and Kytö (1997) indicate, this change took place in the Modern English period. In this section, based on the data shown in Anderwald (2014), the following nine verbs in this category are selected and retrieved, with the total frequency in parenthesis: *arrive* (82), *become* (266), *come* (965), *depart* (65), *enter* (145), *fall* (178), *flee* (29) *go* (814), *rise* (54).² The result is shown in Table 2 and Figure 2 below.

Table 2: The choice of auxiliaries with mutative verbs

Date	<i>be</i>	<i>have</i>	<i>have</i> (%)
1500	22	8	27
1525	175	21	11
1550	72	30	29
1575	110	33	23
1600	147	48	25
1625	69	9	12
1650	82	33	29
1675	121	48	28
1700	97	38	28
1725	81	43	35
1750	151	47	24
1775	138	79	36
1800	50	48	49
1825	67	156	70
1850	31	148	83
1875	109	149	58
1900	15	123	89



Figure 2: The occurrence rate of *have* with mutative verbs

From 1650 onward, the occurrence rate of *have* used in the perfect increased, and reached almost 90% in 1900, within a 250-year development window.

3.3 The development of the progressive (*be -ing*) form

The verbal combination *be + -ing*, often called the progressive, has been much investigated, as in the work of Mossé (1938), or Scheffer (1975). This section deals with the combination which occurred in Middle English and Modern English, as in (4).³

- (4) Our Lord *was penchand* on vs (CMEARLPS-M2,142.6203)
 as the sayd duke of Orleau~ce *was goyng* towarde his lodging
 (FABYAN-E1-P1,560.91)
 that ye with our hole armye *was coming* to the rescue
 (HENRY-1520-E1-H,1.1,236,E.5)
 so they *be speakinge* (ASCH-E1-H,3R.60)
 signs *are* not *wanting* that equally good bulbs can be grown in several places
 in the United Kingdom. (WEATHERS-1913-1,1,3.50)

Table 3 below shows the development of the progressive over time, with raw frequency and its normalized frequency.

Table 3: The development of the progressive form

Date	Frequency	Per Million	Date	Frequency	Per Million
1150	2	172	1550	57	250
1175	-	-	1575	52	225
1200	0	0	1600	75	288
1225	66	331	1625	47	447
1250	-	-	1650	117	597
1275	0	0	1675	190	598
1300	-	-	1700	180	730
1325	7	152	1725	350	1,153
1350	43	967	1750	337	1,029
1375	13	79	1775	485	1,172
1400	34	275	1800	329	1,333
1425	30	201	1825	767	1,892
1450	74	288	1850	568	1,747
1475	15	268	1875	441	1,423
1500	39	191	1900	525	2,502
1525	58	240			

Figure 3 shows the development of the progressive with normalized frequency.

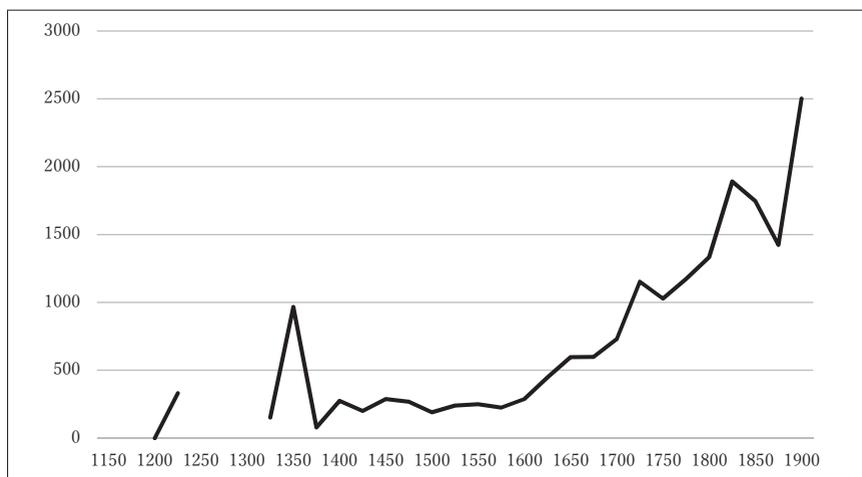


Figure 3: The development of the progressive form

The table and figure indicate that though it was used at an exceptionally high rate in 1350, the construction was used at a low rate through Middle English to the beginning of early Modern English.⁴ The construction was favorably used from 1600 onwards. The characteristics of the combination of *be* with the present participle are under much

debate, as seen in Denison (1993) or Kranich (2015). However, close observations reveal different perspectives. Based on the procedures taken in Petr  (2015), types of common verbs in each period will be shown in Table 4 below.

Table 4: Frequent verb in each period (normalized frequency per one million)

1600		1625		1650		1675	
go	61	want	38	want	41	go	79
come	42	ramble	19	go	41	come	31
remain	12	make	19	come	31	want	31
lean	8	come	19	speak	26	assist	22
baptize	8	ride	19	pass	15	aid	19
want	8	go	19	say	15	make	13
speak	4	mend	10	act	10	pass	9
provide	4	shoot	10	prepare	10	sit	9
live	4	call	10	fly	10	do	9
concern	4	dance	10	do	10	talk	9

1700		1725		1750		1775	
go	97	go	119	go	128	go	128
do	36	come	43	come	61	come	44
speak	24	endeavour	40	sit	40	speak	44
sit	20	speak	33	speak	24	do	29
come	20	do	33	walk	24	ask	24
talk	16	fire	30	do	21	sit	22
want	16	make	30	read	18	live	22
learn	12	stand	26	take	18	say	19
strive	12	bear	26	stand	18	make	19
carry	12	walk	20	pass	15	give	17

In certain periods types of common verbs fluctuated. Around 1625 or 1650, the verb *want* was used quite frequently. This verb does not meet the criterion of the progressive construction in Present-Day English, which indicates that the construction was not yet developed in these periods.⁵ In contrast to *want*, the verb *go* became frequently used from 1675. Its occurrence almost doubled from that time, compared to that of 1650, maintaining the first rank since then. It can be assumed that the actual development began from 1650 or later. From 1675, the most frequent verb was *go*, which is a common verb in Present-Day English. This development indicates that the development of the progressive spans 250 years.

3.4 The emergence of the gerund complement

This section deals with the gerund complement, i.e. the combination of a verb followed by the *-ing* form of a verb. Types of verbs will be selected in the following way: the verbs in this pattern which occur frequently in the corpora, and whose overall frequency is greater than ten times. Here the following types of verbs are dealt with: *avoid* (41), *begin* (12), *continue* (33), *decline* (11), *forbear* (28), *help* (77), *intend* (11), *keep* (16), *mind* (19), *prevent* (27), *propose* (10), *recollect* (26), *remember* (57).⁶

- (5) The business is, he wou'd feign himself Mad, to *avoid going* to Gaol.
(VANBR-E3-P2,74.587)
- thes alsoe were charged to *forbeare treating* of controversyes in Religion
(HAYWARD-E2-H,5.25)
- I could not *help talking* to her. (BRONTE-1848-2,2,290.431)
- Do you *remember being* in his employment on the 8th of September, 1794?
(CROSSFIELD-1796-2,35.10)

The occurrence rate is calculated by dividing the number of the occurrence which is followed by the gerund complement by the number of all the occurrences of the particular verb under investigation, shown in Table 5.

Table 5: The emergence of the gerund complement

Date	V <i>-ing</i>	Verb Occurrence	V <i>-ing</i> (%)
1500	0	230	0
1525	1	456	0.22
1550	1	467	0.21
1575	0	395	0
1600	4	562	0.71
1625	0	291	0
1650	6	399	1.5
1675	6	880	0.68
1700	19	688	2.76
1725	44	888	4.95
1750	27	823	3.28
1775	48	1,182	4.06
1800	33	638	5.17
1825	39	955	4.08
1850	20	745	2.68
1875	18	616	2.92
1900	25	495	5.05



Figure 4: The development of the gerund complement

Table 5 and Figure 4 clearly indicate that the gerund complement rapidly developed from 1625 and reached a high frequency in 1900, which indicates that the development required 275 years.⁷

3.5 The replacement of the genitive case by *of*

In Old English, a noun can be modified by another noun in the genitive case as in (6), whether the genitive precedes or follows the noun. This modification is still possible by use of the apostrophe, the genitive marker. After the decay of the inflectional system in nouns, genitive modification is, in most cases, replaced by the *of* phrase, as in (7).

- (6) Seo sawul soðlice is *þæs lichoman lif* (coelive,ÆLS_[Christmas]:143.114)
the soul truly is the bod's life
'the soul truly is the body's life'
Nu bið ælc mann gefullod on **naman** *þære halgan þrynnysse*
Now is each man baptized on name the holy Trinity(Gen)
'Now each man is baptized on name of the holy Trinity'
(coelive,ÆLS[Ash_Wed]:141.2778)
and atte *yeres ende* they loke after the number or after the price
(CMAELR4,2.31)
- (7) Ic hire bead gymmas and **gyrlan** *of golde* (coelive,ÆLS[Agatha]:35.2031)
I her offered jewels and dress of gold
'I offered her jewels and dress of gold'

Haue also **reuerence** *of* *be gode angel*

(CMAELR3,29.80)

Table 6 below shows the chronological replacement of the morphological case by the periphrastic *of*.

Table 6: The replacement of the genitive case by *of*

Date	Genitive	<i>of</i>	<i>of</i> (%)
800	39	0	0
825	-	-	-
850	89	0	0
875	-	-	-
900	3,773	36	0.9
925	-	-	-
950	5,781	34	0.6
975	-	-	-
1000	11,383	134	1.2
1025	63	0	0
1050	742	28	3.6
1075	719	3	0.4
1100	225	4	1.7
1125	348	145	29.4
1150	448	120	29.4
1175	-	-	-
1200	827	0	0
1225	1,422	1,129	44.3
1250	-	-	-
1275	11	68	86.1
1300	-	-	-
1325	74	920	92.6
1350	59	1,214	95.4
1375	352	5,271	93.7
1400	606	2,767	82
1425	210	3,927	94.9
1450	308	5,583	94.8
1475	64	2,063	97
1500	244	1,160	82.6

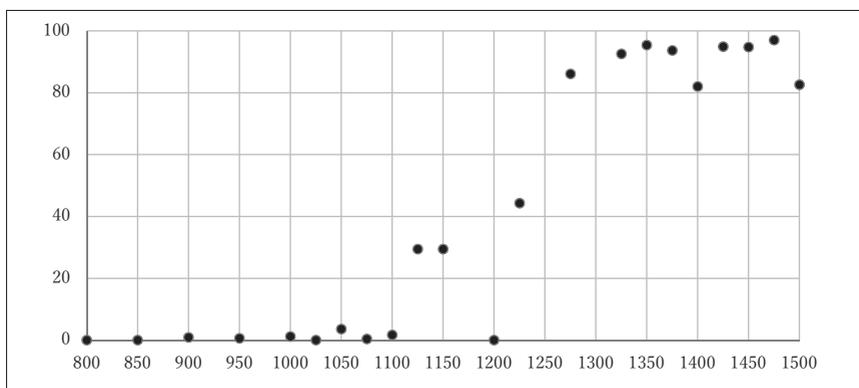


Figure 5: The replacement of the genitive case by *of*

The result indicates that the inflectional genitive has been rapidly replaced by the periphrastic phrase since 1100, and the change was completed in 1300, meaning it had a 200-year span.

3.6 Use of the prepositional adverb *thereof*

In Present-Day English, the use of adverbs which are combined with prepositions, as in *thereof*, *herein*, or *thereby*, is restricted to formal registers. In Middle English or Modern English, however, the prepositional adverb *thereof* was more frequent, was semantically similar to the prepositional phrases ‘of it,’ and “was a frequent alternative” (Nevalainen, 2006: 82) as in (8) and (9).

- (8) the lorde *thereof* was an olde man (CMMALORY-M4,638.3848)
 The lust *thereof* is full of uneasiness; the sating, of repentance
 (BOETHJA-1897-2,90.642)
- (9) a parte *of it* helde ayenst Arthur (CMMALORY-M4,11.318)
 Thou must beg from those who have the giving *of it*
 (BOETHJA-1897-2,92.661)

This section concentrates on the use of *thereof* and its equivalent *of it* in the corpora. Table 7 below shows their occurrences together with frequency per million words.

Table 7: *Thereof* against *of it*

Date	<i>thereof</i>	Per Million	<i>of it</i>	Per Million
1200	0	0	1	0.1
1225	41	2.1	1	0.1
1250	-	-	-	-
1275	0	0	0	0
1300	-	-	-	-
1325	3	0.7	0	0
1350	0	0	3	0.7
1375	45	2.7	17	1.0
1400	24	1.9	8	0.6
1425	26	1.7	27	1.8
1450	37	1.4	10	0.4
1475	5	0.9	5	0.9
1500	71	3.5	27	1.3
1525	93	3.9	196	8.1
1550	127	5.6	212	9.3
1575	132	5.7	116	5.0
1600	130	5.0	156	6.0
1625	17	1.6	170	16.2
1650	42	2.1	352	18.0
1675	94	3.0	500	15.7
1700	36	1.5	174	7.1
1725	55	1.8	190	6.3
1750	15	0.5	162	4.9
1775	22	0.5	194	4.7
1800	28	1.1	126	5.1
1825	16	0.4	195	4.8
1850	13	0.4	120	3.7
1875	40	1.3	115	3.7
1900	0	0.0	96	4.6

Figure 6: *Thereof* against *of it*

Date	- Determiner	+Determiner	+Determiner (%)
1125	3,458	3,415	50
1150	2,148	2,902	57
1175	-	-	-
1200	2,987	3,264	52
1225	9,805	11,763	55
1250	-	-	-
1275	150	260	63
1300	-	-	-
1325	2,686	3,155	54
1350	1,933	3,260	63
1375	8,177	13,079	62
1400	5,375	9,311	63
1425	6,630	11,626	64
1450	9,928	22,099	69
1475	2,035	5,590	73
1500	6,321	15,033	70
1525	7,272	20,173	74
1550	9,747	22,088	69
1575	8,533	19,559	70
1600	7,793	21,894	74
1625	3,891	8,589	69
1650	6,143	16,155	72
1675	10,458	26,122	71
1700	7,443	22,026	75
1725	10,451	27,655	73
1750	10,786	29,020	73
1775	14,759	37,456	72
1800	9,851	23,018	70
1825	12,368	37,838	75
1850	10,998	29,529	73
1875	11,180	26,987	71
1900	7,528	18,513	71

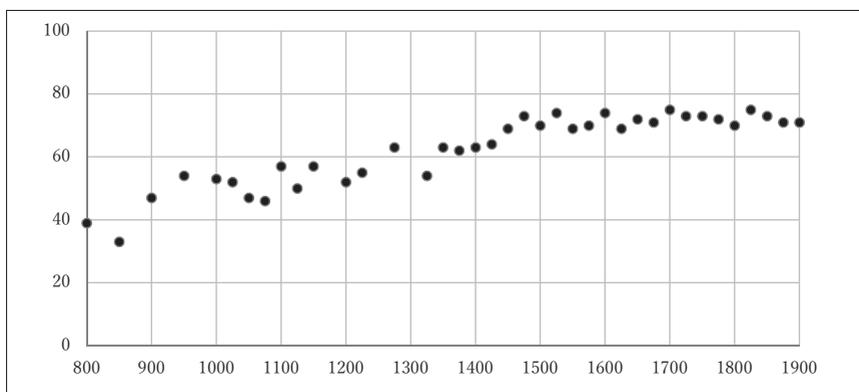


Figure 7: Use of a determiner

The table and figure indicate the use of determiners gradually rose and reached about 70% around 1500. Nevertheless, it never reached 100%. This is because an uncountable noun without a determiner is still allowed in Present-Day English. If this property is taken into account, it is reasonable to claim that the growth of determiner usage lasted 700 years.

3.8 The decline of the double determiner

Two members of the determiner category (i.e. *a, the, this, that, my, you*, etc.) are used in juxtaposition in the Modern English period, as in (12). In this section this type will be called a double determiner. It is no longer allowed in Present-Day English, where the counterparts are formed by one determiner followed by a possessive pronoun like *mine*, or *his*, as in (13).

- (12) Y biseche, for3yue thou the synne of *this thi* puple (CMOTEST-M3,14,1N.646)
 Happy art thou, my scholar, in *this thy* conviction; (BOETHJA-1897-1,98.63)
- (13) & praede him to heue *a* sone of *his* at þe fountston. (CMBRUT3-M3,114.3467)
 Oh, *that* stupid old joke of *yours* about me! (SHAW-1903-2,30.852)

Table 9 and Figure 8 below show the distribution of both types over time. The rate of the double determiner is calculated by dividing the occurrence of the double determiner by the total occurrence of both types.

Table 9: The decline of the double determiner

Date	<i>this _ of mine</i>	<i>this my _</i>	<i>this my (%)</i>
1325	1	0	0
1350	0	1	100
1375	2	1	33
1400	1	0	0
1425	2	1	33
1450	11	2	15
1475	0	2	100
1500	9	18	67
1525	8	29	78
1550	16	33	67
1575	16	24	60
1600	11	30	73
1625	10	17	63
1650	9	16	64
1675	23	8	26
1700	25	0	0
1725	24	0	0
1750	24	1	4
1775	14	2	13
1800	9	0	0
1825	19	2	10
1850	26	1	4
1875	16	2	11
1900	22	0	0

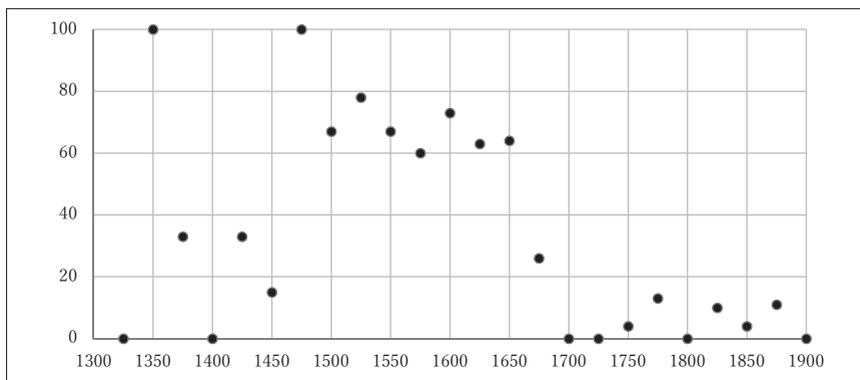


Figure 8: The decline of the double determiner

The double determiner began to be frequent in 1450, and was more frequent than its equivalent around 1500. Then its use became obsolete around 1700. The construction went out of use 250 years later.

3.9 Double comparative and superlative

In certain periods of Modern English, the inflected comparative or superlative can be accompanied by the periphrasis *more* or *most*, as in (14) and (15). These types will be called the double comparative or the double superlative here.

(14) Therefore thou arte *more harder* than ony stone (CMMALORY-M4,656.4474)
For euery vertue that is gadred togyder is *more stronger*.
(FISHER-E1-P1,324.20)

(15) And they welcommed them in the *most faythfullyst* wyse
(CMMALORY-M4,16.481)
Sir, in the *most hartyest* wyse that I can, I recomend me unto you.
(IPLUMPT-E1-H,198.4)

In this section, the focus will be on the beginning of early Modern English, where the duplicated construction, apparent with the double superlative, abounded, as shown in Table 10 and Figure 9 below.⁸ The table shows the raw frequency together with the rate of the double comparative and the double superlative, by dividing the occurrences of the double comparative or superlative by all the occurrences of the comparative or superlative in each period.

Table 10: Double comparatives and superlatives

Date	Double Comparative	(%)	Double Superlative	(%)
1150	0	0	0	0
1175	-	-	-	-
1200	1	0.8	0	0
1225	0	0	0	0
1250	-	-	-	-
1275	0	0	0	0
1300	-	-	-	-
1325	14	5.7	0	0

Date	Double Comparative	(%)	Double Superlative	(%)
1350	0	0	0	0
1375	1	0.2	0	0
1400	4	1.1	2	3.8
1425	3	0.5	0	0
1450	8	0.8	9	9
1475	4	2.4	4	5.8
1500	5	0.6	11	13.1
1525	13	1.4	13	10.2
1550	6	0.6	2	1.2
1575	1	0.1	2	1.7
1600	2	0.2	1	1.1
1625	1	0.2	3	3.9
1650	1	0.1	0	0
1675	0	0	0	0
1700	0	0	0	0

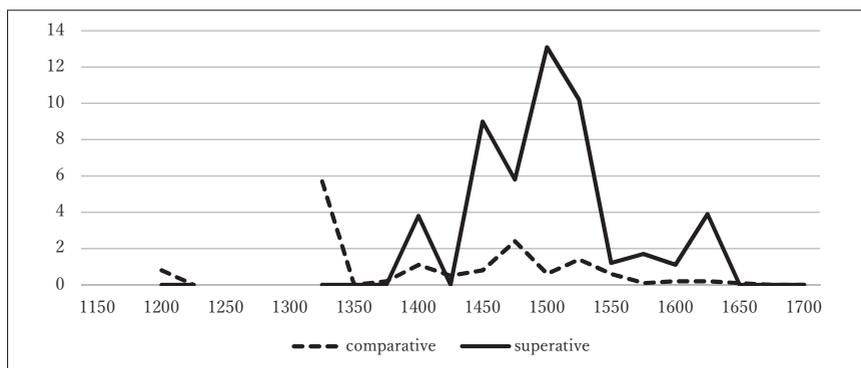


Figure 9: Double comparatives and superlatives

The double superlative began to be more frequent around 1425. One hundred years later, its use began to decline. This decline indicates that the double superlative proliferated in a time span of about 100 years.

3.10 Discussions

The preceding sections showed nine linguistic changes over time. The changes can be summarized as in Table 11 below.

Table 11: Changes of linguistic items

Linguistic Items	Beginning	End	Length
1 Personalization of <i>like</i>	1400	1600	200
2 The choice of auxiliary verbs with mutative verbs	1650	1900	250
3 The development of the progressive form	1650	1900	250
4 The emergence of the gerund complement	1625	1900	275
5 Genitive forms replaced by <i>of</i> periphrasis	1100	1300	200
6 <i>thereof</i> against <i>of it</i>	1475	1625	150
7 Obligatory use of the determiner	800	1500	700
8 The decline of double determiners	1450	1700	250
9 Double comparatives and superlatives	1425	1550	125

The linguistic items (1) to (5) in the table completed their changes over about 200 or 300 years, within similar time spans. These developments reached their goals in 200 or 300 years, giving rise to a new syntactic structure. The linguistic items (6), (8), and (9) showed fluctuations for about 125 or 250 years and then resumed their initial state. It can be stated that the changes which did not reach completion maintained their changes for only 60 or in some cases 150 years. Then the new patterns began to decline, ceasing their changes 125 or 250 years later and becoming obsolete. It is possible to discern the differences of the time span between the completed changes and the uncompleted ones. One could assume that a syntactic change requires a certain duration of time to become complete, which is about 200 or 300 years. This could imply that linguistic change shares a common property.

The change (7), the obligatory use of the determiner, seems quite different from the others. Its beginning was not clear, since the distinction between the determiner and the demonstrative in OE is not clear. However, the rate of the use began to rise and became stable 700 years later. Compared with other productive changes mentioned in this study, this change required as long as 700 years. One may reasonably consider that the transition of this development involves a different property in nature. If the successful shifts involve structural changes, whether they are a tense-aspect distinction as well as the paradigm in the verb, or the inflectional decay followed by the reformation of the noun system, the property concerning the use of the determiner could not be subsumed under the same criteria. It is often assumed that the development of the determiner is associated with the emergence of a functional category in a theoretical perspective, implying a structural or syntactic change.

However, the structural change in the determiner required a different length of time than the other structural changes. It is plausible to speculate that the change in the determiner was driven by a different process.

4. Conclusion

This study examined the time length of changes which took place in the history of English. The linguistic items examined here include the impersonal verb *like*, the *have*-perfect in mutative verbs, the progressive form, gerund complement, genitive case, the use of the prepositional adverb *thereof*, the use of the determiner, the double determiner, and the double comparative and superlative. The data were retrieved from PPCME2, PPCEME, PPCMBE2, and YCOE. Five syntactic changes continued for 200 or 300 years and reached their final stage. The development lasted for about 200 or 300 years. On the contrary, when it declined after about 100 or so years of increase, the change did not remain as a standard linguistic form in English. This result indicates that any change requires a certain duration of time and one can further hypothesize that the length of time required for changes is approximately identical. Looking back at the data concerning the auxiliary *do* provided by Ellegård (1953), the pattern, noticeable in negative questions, started in about 1475 and reached its completion in about 1700, with about a 200 year-time span.

Due to the paucity of text data in the corpora, some of the occurrences provided rather unreliable results, especially in Middle English. The result, however, clearly indicates that there is a similar trend in linguistic development over time. It might be suggested that other fluctuations which became extinct in Present-Day English can be examined under this procedure. It is also possible to extend this research scheme into other domains such as morphological or semantical changes in the history of English.

Acknowledgements

This paper is based on the presentation given at the 44th Conference organized by Japan Association for English Corpus Studies, October 6, 2018. I would like to express my gratitude to three anonymous reviewers for their helpful comments and suggestions, as well as to my colleague Myles K. Chilton for suggesting stylistic improvements.

Notes

1. A narrower segmentation of the corpus causes an uneven distribution of texts into each subperiod. In the Old English period, there are no texts in 825, 875, 925, and 975. On the contrary, the subperiod 1000 contains the largest number of tokens, about 677,000 words, which accounts for approximately half of the corpus. In the Middle English period, no texts are available in 1175, 1250, and 1300. Some segments in Middle English consist of only one text: 1125, 1200, 1275, 1325, and 1350. In the tables below, the subperiod with no token is indicated by “-.”
2. The passive form of the pseudo active phrase *arrive at* is excluded in the results.
3. In what follows, the pattern *be going to*, and passive progressives, are excluded in frequency. Though it is formally similar, the combination of *be going to* does not have exactly the same function as the progressive. As for passive forms, the combination could be subsumed under the verb *be*.
4. The occurrence rate in 1350 is exceptionally high. This segment consists only of the Bible. Since only one type of text is available, this segment may not show a representative property in this period.
5. The common verbs construed as the progressive form in Middle English are *(be)thinken* ‘think’ (32), followed by *wonen* ‘live’ (21), *comen* ‘come’ (16), and *dwellen* ‘remain’ (15).
6. Verbs of inert perception are excluded. In the corpora we have the following occurrences of these verbs: *behold* (11), *find* (30), *hear* (58), *leave* (19), *see* (294).
The section not only deals with the combination of verbs immediately followed by the gerund, but also the combination of verbs followed by the gerund accompanied by its subject, seen in (i).

(i) Do you remember **Mr. Hunt** *stating* any thing

(WATSON-1817-1,1,122.1139)

7. It is not clear that the development ceased in 1900. Rather its development seems to be still underway. For details, see Iyeiri (2010). However, the rate of the gerund complement in Present-Day English was obtained under the same scheme from the *British National Corpus*, the 20th-century English corpus. The appearance rate of the verbal construction investigated is 5.6%. This is a close approximation of the result in 19th-century English.
8. This pattern revived in late Modern English. For details, see Claridge (2006). The rate of the pattern in late Modern English is only 0.5 % (by the calculation taken here). The pattern far more frequently occurred in early Modern English than in late Modern English.

Corpora

- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. (2004) *The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)*. Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, release 3 (<http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCEME-RELEASE-3>).
- Kroch, Anthony, Beatrice Santorini, and Ariel Diertani. (2016) *The Penn Parsed Corpus of Modern British English (PPCMBE2)*. Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, release 1 (<http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1>).
- Kroch, Anthony, and Ann Taylor. (2000) *The Penn-Helsinki Parsed Corpus of Middle English (PPCME2)*. Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, release 4 (<http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCME2-RELEASE-4>).
- Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths (2003) *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*, University of York, York.

References

- Anderwald, L. (2014) “The decline of the BE-perfect, linguistic relativity, and grammar writing in the nineteenth century.” In Hundt, Marianne (ed.), *Late Modern English Syntax*. Cambridge: Cambridge University Press, pp. 13–37.
- Claridge, C. (2006) “‘With the most superlative felicitée’: Functions of the Superlative in 19th-Century English.” In Dalton-Puffer, Christiane, Dieter Kastovsky, Nikolaus Ritt and Herbert Schendl (eds.), *Syntax, Style and Grammatical Norms: English from 1500–2000*. Bern: Peter Lang, pp. 73–89.
- Denison, D. (1993) *English Historical Syntax*. London: Longman.
- Ellegård, A. (1953) *The Auxiliary ‘Do’: The Establishment and Regulation of Its Use in English*. Stockholm: Almqvist & Wiksell.
- Fries, C. C. (1940) “On the Development of the Structural Use of Word-Order in Modern English.” *Language* 16.3:199–208.
- Iyeiri, Y. (2010) *Verbs of Implicit Negation and their Complements in the History of English*. Amsterdam: John Benjamins.
- Jespersen, O. (1909–1949) *A Modern English Grammar on Historical Principles*. London: George Allen & Unwin.
- Kranich, S. (2015) *The Progressive in Modern English: A Corpus-Based Study of Grammaticalization and Related Changes*. Amsterdam: Rodopi.
- Kytö, M. (1997) “Be/have + past participle: The choice of the auxiliary with intransitives from Late Middle English to Modern English.” In Rissanen, M., M. Kytö and K. Heikkonen (eds.), *English in Transition: Corpus-based Studies in Linguistic Variation and Genre Styles*. Berlin: Mouton de Gruyter, pp. 17–85.

- Mossé, F. (1938) *Histoire de la Forme Périphrastique être+participle présent en Germanique*. Paris: Librairie C. Klincksieck. [Translated by Hiroshi Takahashi (1993) 『ゲルマン語・英語迂言形の歴史』 東京：青土社]
- Nevalainen, T. (2006) *An Introduction to Early Modern English*. Oxford: Oxford University Press.
- Petrè, P. (2015) “Grammaticalization by changing co-text frequencies, or why [BE Ving] became the ‘progressive’.” *English Language and Linguistics* 20.1: 31–54.
- Rydén, M. (1991) “The *be/have* Variation with Intransitives in its Crucial Phases.” In Kastovsky, D. (ed.), *Historical English Syntax*. Berlin: De Gruyter, pp. 343–354.
- Rydén, M. and S. Brorström. (1987) *The be/have Variation with Intransitives in English, with Special Reference to the Late Modern English Period*. Stockholm: Almqvist and Wiksell.
- Scheffer, J. (1975) *The Progressive in English*. Amsterdam: North-Holland Publishing Company.
- Traugott, E. C. (1992) “Syntax.” In Hogg, Richard H. (ed.), *The Cambridge History of the English Language I*. Cambridge: Cambridge University Press, pp. 168–289.
- Visser, F. Th. (1963–1973) *An Historical Syntax of the English Language*. Leiden: E. J. Brill.

(塚本 聡 日本大学 Email:tukamoto@chs.nihon-u.ac.jp)

英語コーパス学会 第44回大会

■第1日目

ワークショップ受付：2号館1階 211教室前にて9:30受付開始
 ワークショップ参加費：会員無料。非会員2,000円（当日会員としての大会参加費、2日間共通）。

ワークショップ1【言語研究のための Word2Vec 入門】

会場：東京理科大学（神楽坂キャンパス）2号館2階 223教室
 日時：10月6日（土）10:00-11:00
 講師：内田 諭（九州大学）

ワークショップ2【コーパス研究の作法】

会場：東京理科大学（神楽坂キャンパス）2号館2階 223教室
 日時：10月6日（土）11:00-12:00
 講師：井上 永幸（広島大学）

日 時	2018年10月6日（土）	
受付開始	12:00（2号館1階 211教室前）	
開 会 式	13:00（2号館1階 212教室）	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		渡辺 一之（東京理科大学副学長）
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：2号館1階 212教室）〉

司 会 柴崎礼士郎（明治大学）

研究発表1 14:00-14:30

英語史的コーパスによる言語変化速度の測定 塚本 聡（日本大学）

研究発表2 14:35-15:05

アメリカ英語における分離不定詞使用に関する研究：“splitter”に焦点を当てた分析
 福本 広光（大阪大学大学院生）

研究発表3 15:10-15:40

コーパスを利用した18・19世紀オーストラリア文学作品における

自動詞完了形助動詞に関する研究

守家 輝 (京都大学大学院生)

〈研究発表第2セッション (場所: 2号館2階 223教室)〉

司会 水本 篤 (関西大学)

研究発表1 14:00-14:30

(キャンセル)

研究発表2 14:35-15:05

Using Corpora to Examine Lecturing Styles in American and Japanese University

Engineering Courses

Judy Noguchi (Kobe Gakuin University)

Kazuko Tojo (Osaka Jogakuin University)

Nilson Kunioshi (Waseda University)

研究発表3 15:10-15:40

Collaborative Texts under a Stylometric Microscope: Investigating Texts of Mixed Authorship

Tomoji Tabata (Osaka University)

〈休憩 15:40-16:10〉

〈シンポジウム: 16:10-17:40 (場所: 2号館1階 212教室)〉

Issues on multi-word units (MWUs) and collocation

Chair: Laurence Anthony (Waseda University)

On the Creation of a Large-Scale Multi-Word Unit Resource for Learners of English
for Academic Purposes

James Rogers (Meijo University)

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

Atsushi Mizumoto (Kansai University)

Collocational patterns beyond word pairs

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

〈懇親会 18:00-20:00〉

■第2日目

日 時 2018年10月7日（日）
 受付開始 9:30（2号館1階 211教室前）

〈研究発表第3セッション（場所：2号館1階 212教室）〉

司 会 久保田俊彦（明治大学）

研究発表1 10:00-10:30

Charles Dickens の *The Mystery of Edwin Drood* の, “Thomas Power James” による続編の
 「著者推定」-テキスト全体・地の文・発話の3種類の語彙による推定の比較・評価
 後藤 克己（中部大学大学院生）

研究発表2 10:35-11:05

（キャンセル）

研究発表3 11:10-11:40

トピックモデルを用いた Agatha Christie 作品へのアプローチ

土村 成美（大阪大学大学院生）

〈研究発表第4セッション（場所：2号館2階 223教室）〉

司 会 能登原祥之（同志社大学）

研究発表1 10:00-10:30

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

西垣知佳子（千葉大学）

赤瀬川史朗（Lago 言語研究所）

中條 清美（日本大学）

研究発表2 10:35-11:05

英日バイリンガルエッセイコーパスに見るコロケーションの比喩的な意味拡張

鎌倉 義士（愛知大学）

研究発表3 11:10-11:40

英語学習者コーパス構築のためのタスク設計：特定の文法項目抽出に向けて

工藤 洋路（玉川大学）

内田 諭（九州大学）

〈休 憩 11:40-13:00〉

〈研究発表第5セッション（場所：2号館1階 212教室）〉

司会 秋山 孝信（日本大学）

研究発表1 13:00-13:30

アカデミックライティングにおけるヘッジの活用：研究論文における Discussion の
コーパス分析 中谷 安男（法政大学）

研究発表2 13:35-14:05

強意語的機能を持つ罵倒語の進化特性について 新井 洋一（中央大学）

〈研究発表第6セッション（場所：2号館2階 223教室）〉

司会 阿部真理子（中央大学）

研究発表1 13:00-13:30

Reliability and Replicability of Annotation Schemes for Learner Corpora
Aika Miura（Tokyo University of Agriculture）

研究発表2 13:35-14:05

Searching for grammatical items as criterial features of CEFR levels in spoken and written
learner corpora: Using the CEFR-J Grammar Profile
Yukio Tono（Tokyo University of Foreign Studies）
Yasutake Ishii（Seijo University）

〈休憩 14:05-14:30〉

〈講演 14:30-16:00（2号館1階 212教室）〉

司会 投野由紀夫（東京外国語大学）

Measures of Productivity and Lexical Diversity

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

閉会式 16:00（2号館1階 212教室）

閉会の辞

■10月6日（土）

【ワークショップ1】

言語研究のための Word2Vec 入門

内田 論（九州大学）

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

本ワークショップでは、多くの自然言語処理の研究で用いられるようになった **word embedding**（単語埋め込み）について基礎的な概念を説明し、その中でも特に広く利用されている **Word2Vec** について概観する。**word embedding** は単語の意味をベクトル（数字の集合）として扱う技法で、意味の似た単語は似た文脈で用いられるという分布仮説が背景にある。単語の意味をベクトルで表すことで、単語と単語の距離を測定することが可能となり、類義語の抽出やベクトルを用いた意味の演算が可能となる。これらの性質を用いることで品詞解析や機械翻訳、語義曖昧性解消などの精度が向上することが報告されており、自然言語処理において基幹的な技術の一つとなりつつある。**Word2Vec** は Google の研究チームが開発した **word embedding** のアプリケーションの一つで（cf. Mikolov et al. 2013）、Python や TensorFlow などの広く使われているプラットフォームを通して比較的容易にかつ高速に実行することができるものである。本ワークショップではウェブのインタフェース（Python CGI）を用意し、**Word2Vec** を実際に動作するところを示す。例えば、約 1 億語のコーパスを用いて作成したモデルを用いて **speak** の類義語（ベクトル間の距離が近いもの）を検索すると、**talk, respond, communicate** などの意味が近いと考えられるものや、**listen, hear** などの知覚を表す動詞などがリストされる。このような結果が言語研究にはどのように応用できるかを議論し、**Word2Vec** を用いた新たなコーパス研究の可能性について探る。

【ワークショップ2】

コーパス研究の作法

井上 永幸（広島大学）

比較的簡単に大規模コーパスにアクセスできるようになったり、膨大なインターネット上のデータをコーパス代わりに利用する機会が増えるなど、コーパス研究のための環境変化には目を見張るものがある。本ワークショップでは、特にコーパス研究初心者が遭遇しそうな状況を想定しながら、問題解決の際の糸口を提案してゆく。

大規模コーパスの利点は、それほど頻度の高くない語句について、小さなコーパスでは特徴として認識できなかった形が特徴として認識できるようになることであろう。たとえば、*enough as it is* は、Brown, LOB, Frown, FLOB を合わせた400万語のコーパスでは1例しか出てこないが、4億6230万語の *WordbanksOnline* では100例出てきて、その定型表現的性質を垣間見せてくれる。また、大規模コーパスを使った研究の醍醐味を見せてくれるのは各種統計値であろう。MI-score は、低頻度のものにも反応するため、その取っ付きにくさから敬遠されがちであるが、逆にキーワードと連想関係の強い語を高頻度のものから低頻度のものに至るまで焦点を当ててくれるので、キーワードの特性を見事にあぶり出してくれるのである。無論、MI-score が高くて頻度も高い方が信頼度も高まるので、そういった語から優先的にアプローチするのが正攻法であろう。MI-score の「機微」にふれてみたい。

インターネット上にあふれるデータをコーパスとして利用する方法も注目を集めている。日々増大するデータを言語資料として利用しない手はないであろう。ただ、Mark Davies による *BYU corpora* で提供されている140億語から成る *iWeb* で複数形の “*informations*” を検索すると8,000を越える例が出てくることが分かる。英語を地球規模で研究するのであればともかく、少なくとも教育現場で使う英語を吟味するには適さないであろう。ウェブデータをコーパスとして用いる際の留意点などにもふれたい。

■10月6日（土）

【研究発表第1セッション】

【研究発表1】

英語史的コーパスによる言語変化速度の測定

塚本 聡（日本大学）

古英語以来、英語には多くの統語的变化が生じている。Denison (1993) が示す通り、動詞にかかわる変化は特に顕著である。Fries (1940) や Ellegård (1953) をはじめ、個別の言語変化についての研究は多数行われている。それらの個別の研究では、各言語変化の開始時期や終了時期、あるいは量的な変化は示されるものの、言語変化に要した時間を横断的に扱った研究は見られない。本研究では、以下に挙げる言語項目の変化速度について、コーパスを使用し、その変化に要した時間を観察し、その変化速度にある一定の規則性がみられるかを検討する。

異なる言語資料を使用することによる差異が生じることが無いよう、本研究では Penn-Helsinki Parsed Corpus (ME2, EME, MBE2) を中心的言語資料とし、YCOEなどを追加的に使用し、下記の言語項目の変化を観察する。

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

具体的には、(1) ME 後期から eModE にかけて非人称動詞 like が人称動詞化した変化、(2) ModE 後期に be+V-en であった変移動詞の助動詞が have+V-en に変化した現象、(3) eModE に be+ing 構文の主要な動詞が dwell などの非有界動詞から有界動詞へと拡大した変化、(4) ModE 後期に、一部の動詞補部に -ing が生じしその頻度を高めた変化、(5) OE から ME にかけて屈折語尾の衰退に伴い前置詞の使用が増加した変化、(6) ほぼ同義となる thereof および of it は一部のレジスターを除き、後者の使用が優勢となる変化、(7) 可算名詞において、a, the, my などの限定詞が必須となる変化、(8) ModE まで生起していた these my words のような限定詞の連鎖が許容されなくなった変化、(9) most easiest のように迂言形を比較級・最上級に使用する現象、を調査対象とする。なお、(8) および (9) は定着しなかった変化である。

調査は Penn-Helsinki Parsed Corpus で解析された統語情報から条件に合致する構造を検索し、その生起数や対立する形態・構造との構成比率の変化を観察することに

より行う。予備的調査では、定着した変化のいくつかは、ほぼ200~250年くらいの時間で変化が完了していることが観察されている。これらの観察から、言語変化は一定の変化速度を有していることを示す。

【研究発表2】

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

福本 広光（大阪大学大学院生）

本研究の目的は、アメリカ英語における分離不定詞の使用の実態について、特に to と原形動詞を「分離」する要素である ‘splitter’ を中心に通時的に考察することである。

分離不定詞は、Rohdenberg (2009), Mitrasca (2009), Mikulova (2011) などの研究により、イギリス英語よりもアメリカ英語において頻繁に確認される構造であることが分かっている。そのことから、アメリカ英語は分離不定詞の主要な使用域であると考えられる。また、分離不定詞に関する通時の研究は Calle-Martin and Miranda Garcia (2009) などを除いて多くはなく、計量的に見るとまだ知られていないことも多いと思われる。本研究では主に COHA, Time Magazine corpus の2種類のアメリカ英語コーパスを用いる。

本研究における RQ は以下の3つである。

1. 分離不定詞は、アメリカ英語においてどのように拡大してきたのか
2. splitter そのものに、経年的な変化というものがあるのか、またどのような変化なのか
3. 分離不定詞が、どのような context で生起しているか

この“splitter”を、例えば (1) での really のような1語 splitter と (2) での all of a sudden のような複数語 splitter に分け、それぞれについて通時的コーパス COHA を用いて量的・質的に経年的変化について分析する。

(1) He seemed *to really want* to find a way to serve. (Time Magazine Corpus)

(2) The states he’s won, the red states, is they’re not likely *to all of a sudden turn* blue in November. (COCA)

具体的には、量的分析として、分離不定詞の200年間の頻度変遷 (Fiction などのジャンル別の頻度変遷を含む) を調査し、さらに各ジャンルでの頻度変化について見ることによって頻度変化のメカニズムを探る。

1語 splitter を伴う分離不定詞の使用頻度は、20世紀前半まで増減を繰り返していたが、20世紀後半、特に1990年以降大きく伸びている。その背景には Fiction と Magazine ジャンル、とりわけ Magazine ジャンルにおける増大がある。(Time Magazine corpus

でも同じような結果が得られた) また、20世紀初頭に分離不定詞は頻度を減少させているが、その背景に当時の分離不定詞のメイン使用域であった Fiction における減少があった。これらの時代の Magazine や Fiction をさらにサブジャンルに下位分類し、当時の使用の実態について考察する。

splitter の経年的変化については、おもに COHA の Fiction ジャンルから分離不定詞として使用された副詞をすべて抽出、集計し、時代ごとに偏って使用されている splitter の存在を観察する。そして長年にわたって使用されてきた副詞 (even, just など) もある一方で、時代ごとに特徴的な副詞 (thus, actually など) もまた存在することを明らかにする。

分離不定詞には、「使用すべき理由」もまた存在する。「曖昧性を解消し、文の意味を正確に伝えるとき」や、「リズムを整え、発音しやすく心地よい響きを与えるとき」など (Crystal (1984:29-30)) である。抽出した例文のうち、どれだけがこれらの理由に則って使用されているのか、コンコーダンスライン全体の意味を解釈しながら考察する。

複数語からなる splitter が使用される場合、そのパターンとして

- (i) at once や at least などの、lexicalize されたもの
- (ii) more strongly や at least passively などの、一方 (一部) の副詞がもう一方の副詞を修飾するもの

のどちらかに区分されるということがほとんどであるということが確認された。

【研究発表3】

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

守家 輝 (京都大学大学院生)

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

Have 完了形の増加現象を研究する上で、各植民地社会で新たな言語変種が成立しつつあった後期近代の特性から鑑みて、地域・社会的要因を検証する余地が多分にある。本研究では特に、19世紀以前のオーストラリア社会における、アイルランド

英語 (IrE) の影響という側面に焦点を当てる。アイルランド移民は19世紀以前オーストラリア入植者の約20%を占め、イングランド移民に次いで多い (Fritz 2007: 22-27)。IrE 由来の統語的特徴は小説・演劇等の会話文でアイルランド人の発話内に頻出するものがある一方、一部はオーストラリア英語 (AusE) 特有の表現として定着した (Burridge and Musgrave 2014: 40-41)。自動詞 *have* 完了形の発展に関する IrE 等言語変種の影響を解明する上で、初期オーストラリア社会を対象とした社会言語学的研究は有意義だと考えられる。

本研究では18-19世紀のオーストラリアで執筆された文献を収録した *Corpus of Oz Early English* (COOEE) と、*AustLit* の2種類のコーパスから抜粋した、18・19世紀の fiction ジャンルに分類されるテキスト約300万語を対象に分析を行う。Rydén and Brorström (1987) を参考に19世紀において *be/have* の交代が確認される自動詞12種を選び、各自動詞の完了形構文中で *have* が選択される比率を計算して、その結果を話者 (著者あるいは発話者である登場人物) の出身地・社会階層という観点から分析する。また同様の分析を ARCHER コーパスに収録された同時代・同ジャンルのイギリス英語 (BrE)、アメリカ英語 (AmE) テキストにも適用し、AusE と比較する。

その結果、AusE、BrE、AmE の三変種間における *have* 完了形使用頻度の推移に差があることが判明し、その変化が確認される時期とオーストラリアへのアイルランド系移民の流入時期との間に相関がある可能性が示唆される。また、同じアイルランド系作家の文章でも、AusE のデータでは *have gone* の使用率が減少傾向にある期間において、BrE では逆に増加傾向である等、移民と本国の作家との間で特徴の差異が認められる。さらに、“an’ I wouldn’t a’ come anigh you” (*AustLit. Dave’s Sweetheart.*) のような会話文中で助動詞の形態が特殊な場合 (*a’ come*) が確認されることなどから、初期のオーストラリア移民が用いた言語変種の特徴が作品の著者によって登場人物が用いる完了形構文においても表現されているという仮説を立て、*have* 完了形の使用率を発話者の出身地・社会階層といったオーストラリア植民地時代における社会的側面との関係から分析し、解明していく。

■10月6日（土）

【研究発表第2セッション】

【研究発表1】

（研究発表第2セッションの研究発表1はキャンセルされました。）

【研究発表2】

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

Judy Noguchi (Kobe Gakuin University)

Kazuko Tojo (Osaka Jogakuin University)

Nilson Kunioshi (Waseda University)

The rapid globalization of society today has accelerated the need for internationalization of higher education in Japan. In 2008, the Japanese government announced the “300000 Foreign Students Plan” Campaign (Ministry of Foreign Affairs, 2010) and as of May 1, 2017, the number of international students studying in Japan had reached 267,042, with 188,384 students being enrolled in institutions of higher education (Japan Student JASSO, 2017). This has led to an increase in the demand for university degree courses which are offered using English as the medium of instruction. With an interest in aiding Japanese instructors faced with delivering lectures in their disciplines in English, especially those in science and engineering, we built OnCAL, a corpus of university lectures for science and engineering courses. We aimed at identifying the pedagogical functions and useful expressions that could be used for delivering lectures in English. The OnCAL concordancing interface allows access to 430 science and engineering lectures given at MIT OpenCourseware (MIT OCW, <http://ocw.mit.edu/index.htm>) and Stanford Engineering Everywhere (SEE, <http://see.stanford.edu/>). This work led us to wonder about how functions that we had uncovered, such as the asking of questions and the proposal of thought experiments to initiate consideration of the lecture contents, would be delivered in comparable lectures at Japanese universities. We therefore began construction of a corpus of Japanese lectures given at four major national and private universities in similar disciplines. At present, there are 104 Japanese lectures and comparing them with the American lectures revealed marked differences in the lecturing styles. In this paper, we focus on questions and their functions in the lectures. In general, the 430 lectures given at American universities included many question-word phrases to elicit student thinking with frequent uses of personal pronouns such as *you* and *we*, indicating more interaction during the class. Question-word phrases with *what* (1706 average per lecture), *how* (697), and *why* (244) were frequently used, in contrast to the Japanese lectures where 何, なん (*what*)

was used only 2.8 times on average per lecture, なん (て), どう, どのよう (*how*) only 2.8 times and 何 (で), なん (で), どう (して), なぜ (*why*) only 0.6 times. The lectures at Japanese universities were presented in a manner suggesting the presence of a top-down authority. Overall, the American lectures tended to be presented in a conversational style with a strong audience orientation while the Japanese lectures were given in a more formal style with an emphasis on content dissemination. Our findings led us to conclude that English-medium instruction may not be successful if a lecture which was originally intended for a Japanese university audience was just simply delivered in the English medium. What is essential for successful instruction in a “globalized” classroom with students from different educational backgrounds is an awareness of differences in lecturing styles and the structuring of lectures in order to reach students who may have different expectations with respect to classroom instruction.

【研究発表3】

Collaborative Texts under a Stylometric Microscope: Investigating Texts of Mixed Authorship

Tomoji Tabata (Osaka University)

The Victorian author Charles Dickens was among the first publishing entrepreneurs to run mass-produced weekly magazines on a successful commercial basis. He employed many “salaried staff writers” (Nayder, 2002), who had to write under anonymity, including Elizabeth Gaskell, Adelaide Anne Proctor among others, in *Household Words* and *All the Year Round*, the journals “conducted by” Dickens (Stone, 1968; Thomas, 1982; Allingham, 2011). On the other hand, Dickens collaborated with his younger contemporary Wilkie Collins on a number of stories, typically for the Christmas Numbers of his journals. While some of their collaborative pieces were written with the assistance of other staff writers, four works are known to have been co-authored by Dickens and Collins alone (Nayder, 2002): *The Frozen Deep* (1857), *The Lazy Tour of Two Idle Apprentices* (1857), *The Perils of Certain English Prisoners* (1857), and *No Thoroughfare* (1867). The four collaborations can be seen as betokening what appears to be a firm presence of Collins, a foothold he had gained, in the Dickens circle by the time he and Dickens launched into the joint works beginning in 1857.

The present study draws on a corpus of Dickens set comprising 22 texts and a Collins set with the same number of texts as a training corpus, with which we compare the style of the collaborative texts. Just as Dickens was a prolific writer, so was Collins. His career spans 38 years including his ‘études’, *Antonina* (1850) and *Rambles Beyond Railways* (1851), which are not included in this study. The set of corpus texts can be transformed into a vector of figures, word-frequencies. Word-frequency profiles make it possible to compare between texts to see a certain set of texts have more frequent access to a particular set of words than others,

while another group of texts may be characterized by consistent avoidance of a particular set of words among others. However, the larger number words the frequency profiles encompass, the greater difficulty they impose upon us when we try to grasp complex interrelationships between a large number of vocabulary items and relationships between texts. This is exactly where we need a statistical technique for visualization. This study employs Support Vector Machine in an effort to attribute a chunk of text in question to the more likely author with rolling chunks of the collaborative text progressively compared with the training corpus.

A rolling chunk is designed to be sensitive to a stylistic shift in texts in order to pinpoint where one author takes over from the other in the text of mixed authorship. For this study, collaborative texts are segmented into equal-sized, partially overlapping chunks. If we specify a ‘chunk size’ of 3,000 and a ‘step size’ of 300, for example, the first chunk of a text contains 1st–3,000th words, the second has 301st–3,300th words, the third 601st–3,600th words, and so forth. The procedure uses the relative frequencies of n most frequent words in the reference collection. Emerging results from this analysis show that it is possible to assign a particular chunk to the more likely author with a 100% accuracy, thus allowing us to locate exactly where authorial takeovers happen in the texts of mixed authorship. The findings from this research show that it is always Dickens who starts joint chapters, setting the keynote of each of the collaborated chapters. A typical pattern is that Dickens runs about one-third to halfway into a chapter before passing over to Collins. Of remarkable interest with respect to the making of collaborations is that the pattern in authorial takeovers can be interpreted as reflecting an “unequal” partnership (Nayder, 2002) between Dickens and Collins, just like one between a master and his disciple.

【シンポジウム】

Issues on multi-word units (MWUs) and collocation

Corpus linguistics research is paying increasing attention to the identification and usage of multi-word units (MWUs). To date, research on MWUs has provided new insights in a wide range of areas, including vocabulary development, learner writing, discourse signaling, and disciplinary variation. However, the precise definition of MWUs has proved elusive, leading to widely different interpretations of what counts as a MWU and how they are used in spoken and written language. One common view has been to regard MWUs as contiguous word units that appear above a threshold frequency and dispersion value in a target corpus. Another common view has been to regard MWUs as contiguous and non-contiguous word units in which the members have a collocation strength above a certain statistical or effect-size threshold. As a result of the current confusion, corpus tools provided widely differing functions for identifying and analyzing MWUs. They also differ greatly in regard to the

measurements of association between candidate members offered to researchers.

In this symposium, MWUs and their relation to the concept of collocation will be discussed from three different perspectives: the creation of high-frequency, pedagogic MWU lists using lemmatized collocates; applications of MWUs in the development of practical learning and teaching tools, and novel methods for automatically identifying non-idiomatic MWU combinations. Following presentations by three influential researchers in this field, the audience will be invited to join the panelists in a discussion of collocation measures and the challenges of identifying and utilizing multi-word units (MWUs).

Chair: Laurence Anthony (Waseda University)
Issues on multi-word units (MWUs) and collocation

To open the symposium, I will briefly review some of the definitions of multi-word units (MWUs) proposed in the literature and summarize the various challenges that researchers face when identifying and utilizing MWUs in linguistic research as well as language learning and teaching resources. As part of this introduction, I will briefly discuss a new software tool designed specifically for MWU identification and analysis.

Panelist 1: James Rogers (Meijo University)
On the Creation of a Large-Scale Multi-Word Unit Resource
for Learners of English for Academic Purposes

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

Panelist 2: Atsushi Mizumoto (Kansai University)
Applying a bundle-move connection approach to the development of
an online writing support tool for research articles

Achieving a high level of English proficiency requires a comprehensive English vocabulary of which multi-word units (MWUs) are a critical component. However, acquiring and using these MWUs poses a formidable challenge for second language users of English. In order to facilitate the learning of these units, various online reference resources based on different types of corpora have been developed in recent years. Also, there is a growing interest in resources that are specifically designed to help learners develop an understanding of MWUs above the level of the sentence. In this talk, I introduce some of the current resources available for accessing MWUs that can help to develop rhetorical competency. Specifically I will focus on a data-driven and theory-based practical writing support tool for research articles (RAs) called AWSuM. This innovative, web-based tool is powered by a combination of rhetorical moves and lexical bundles. It also has an auto-complete feature that suggests the most frequent lexical bundles in a move within an RA section. AWSuM was developed as a proof-of-concept of the bundle-move connection approach. Preliminary user feedback was positive overall, and the writing support tool was found to bring about beneficial effects that genre writing pedagogy explicitly aims to achieve. In light of these findings, the pedagogical implications of the developed tool are discussed, with particular focus on the potential role that it can play in the teaching and learning of technology-enhanced genre writing.

Panelist 3: Stefan Evert (Friedrich-Alexander-University of Erlangen-Nürnberg, Germany)
Collocational patterns beyond word pairs

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

of a MWC can be modelled in terms of nested hypothesis tests, taking into account both significance and association strength. Such nested hypotheses may also involve semantic or morphosyntactic restrictions on the slots, or test whether a larger MWC is composed of overlapping smaller MWC (e.g. “earn good money” from “earn money” + “good money”).

■10月7日（日）

【研究発表第3セッション】

【研究発表1】

Charles Dickens の *The Mystery of Edwin Drood* の,
“Thomas Power James” による続編の「著者推定」

ーテキスト全体・地の文・発話の3種類の語彙による推定の比較・評価

後藤 克己（中部大学大学院生）

米国人 Thomas Power James (以下, James) は, Dickens の遺作となった *The Mystery of Edwin Drood* (以下, *ED*) に続編を加えた「完全版」(Dickens, [James], 1873) を発表し, その続編を “By the Spirit Pen of Charles Dickens, through a Medium.” とアピールしている。後藤 (2017) は地の文の語彙頻度を用いた対応分析結果, およびコロケーションの生起頻度の比較から, *ED* と続編の文体の類似性は低いと結論づけ, このアピールを疑わしいものと評価している。

本研究では, 対応分析とともに文体比較／著者推定に多く用いられている多次元尺度法およびクラスター分析を使用し, あらためて続編の著者が「Dickens の霊」といえるかどうかの推定を試みる。この推定は, 続編が Dickens の霊による自動書記 (automatic writing) で創作されたならば, それは Dickens の文体特徴をもつはず, との考えによる。なお著者推定には, 地の文に生起する語彙頻度が多く用いられる (Hoover, 2001, 2002) が, 一般に発話テキストの除去には, 1重引用符とアポストロフィの区別, 引用符を欠いた自由直接発話の抽出などが必要となり大きな負担となる。そこでテキスト全体の語彙／地の文のみの語彙のそれぞれで分析し, 分析語彙を地の文に限定することの効果を評価する。

コーパスには後藤 (2017) と同様に, *ED* と続編に, 参照用として Dickens の *Our Mutual Friend* (以下, *OMF*) を加えた3作品を用いた。これらを地の文と発話に分離し, それぞれ4～12のサブコーパスに分け Lemma による語彙頻度数を抽出した後, Hoover (2004) を参考に1作品に70%以上偏在する語, 人称代名詞および固有名詞を除外した。このデータをもとにテキスト全体・地の文・発話の3種類のデータを構成した。

分析の結果, 地の文の語彙はもとよりテキスト全体の語彙によっても, さらには発話の語彙によっても, 続編のサブコーパスは *ED*・*OMF* のそれとは別クラスター

を構成しており、続編を Dickens の「遺作」とする James のアピールは疑わしいとの、後藤 (2017) と同じ結論を得た。

なお、多次元尺度法の散布およびクラスター分析の樹形図において、続編と *ED・OMF* との距離は、地の文語彙の場合が最も大きく、テキスト全体、発話の順に減少しており、地の文語彙によった場合の著者推定性能の高さが確認された。なお、本研究ではテキスト全体でも地の文と同様な結果が得られており、推定すべき著者が限定的な場合であれば、テキスト全体の語彙の使用も許容されることを示唆している。

【研究発表2】

(研究発表第3セッションの研究発表2はキャンセルされました。)

【研究発表3】

トピックモデルを用いた Agatha Christie 作品へのアプローチ

土村 成美 (大阪大学大学院生)

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

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

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

Christie 作品と特に関連性が強いトピックを確認すると、第一に挙げられるのは、縮約表現を中心とした口語的表現から成るトピックである。このトピックには *I'm*,

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

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

■10月7日(日)

【研究発表第4セッション】

【研究発表1】

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

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

本研究グループでは、小学校英語において、児童が英語の語彙・文法のルールを発見して学ぶ「データ駆動型学習」(Data-Driven Learning: 以下、DDL) を実践するための、英語 DDL 支援サイトを開発している。本発表では特に、その開発に向けた1) 英文事例参照用コーパスの構築、2) 検索ツールの開発、および3) 作業を進める中で遭遇した課題とその解決方法について報告するものである。

発表者らは、DDL を小学生から大学院生の授業に導入し、その普及に努めている。そのなかで、小学校ではペーパー版 DDL を実践し、言葉への気づきを引き出し、英語の語彙・文法の深い学びの促進に効果をあげてきた。小学生のペーパー版 DDL の実践にあたっては、入門期の学習者を、英語のルールやパターンの発見に導くように精選された英文の提示が不可欠で、英文の長さ、語彙・文法のレベル、内容やトピックなどの難易度に関して格段の配慮が要であり、その作業は容易ではない。そこで

発表者らはそのような例文を集積した「英文作例参照用コーパス」を構築した。

実際の作業では、はじめに、英文用例作成の際に参照するための2,500万語規模の入門期英語を集めたコーパスを作成した(2016~2017年度)。ソースとなるデータは、アメリカの教科書、日本を初めとするアジア諸国の検定教科書、入門期英語教材、Graded Reader、語彙を制限した子ども向けのニュースなどのインターネット上の入門期英語コンテンツなどから精選した。この参照用コーパスのデータは、センテンスに区切り、Stanford POS Tagger を用いて品詞タグ付けを行った。

次に、参照用コーパスを活用するための検索ツールを開発した(2018年度前半)。効率的な用例作成を支援するという課題を解決するために、以下のような工夫を取り入れた。

- (1) 検索結果はセンテンスの短いものから順に表示される
- (2) 検索するセンテンスの長さを語数で指定できる
- (3) 文法項目が検索できるように、表層形だけでなく、レマや品詞を指定した CQL (Corpus Query Language) による検索ができる

(1)、(2)については、参照用コーパスにセンテンス長の情報を付与し、センテンス長とセンテンスの文字列をソートキーとしてコーパスの全用例を並べ替え、重複するセンテンスを排除した。次に、センテンス長が1語から100語までの用例のサブセット(約197万センテンス、約2,234万語)を作成し、作例参照の用途により適した形に再編集した。

検索ツールの開発により、用例作成の効率が大幅に向上した。また、英語教育の専門家ではない小学校英語指導者が用例作成に参加できるようになり、授業で使いたいような英文を DDL 支援サイトに組み込むことができるようになった。

【研究発表2】

英日バイリンガルエッセイコーパスに見るコロケーションの比喩的な意味拡張

鎌倉 義士(愛知大学)

第二言語学習者の習熟度を測る指標としてメタファーを含む比喩表現の使用が注目されている。metaphoric competence (Littlemore, 2001) の概念において、目標言語での比喩表現の理解や運用の度合が学習者の言語の熟達度と関連すると言われる (Littlemore, Krennmayr, Turner & Turner, 2014)。本研究では日本人英語学習者の比喩表現使用の分析に、英語と日本語の両言語でのエッセイをデータとするバイリンガルコーパスを使用した。なぜなら、学習者が英語で表現できなかった比喩も対訳の日本語作文でその内容を確認することが可能であるからである。特にコロケーションに注目し日本人学習者の英作文に確認される基本的な意味から比喩的な意味への拡張について分析を行う。例えば、big, serious, significant, important など重大さを示す形

容詞が **problem (s)** と共起する頻度が高い一方、**complicated, complex** という困難さを示す形容詞や **basic, central** など基礎的な意味を指す形容詞は日本人学習者の英作文ではネイティブよりも頻繁に使用されていない。このように日本人英語学習者のコロケーションの特徴を **problem (s)** との共起に限定して本研究では分析する。

研究課題は以下の3つとなる。1) **problem (s)** と共起する形容詞・名詞・前置詞の分析から日本人英語学習者に特徴的な **problem (s)** の意味拡張ネットワークが存在し、記述することは可能か。2) **problem (s)** の意味拡張ネットワークは単純に対訳語となる「問題」とその共起する日本語の語彙の組み合わせとは異なり、日本人学習者の英語独自の意味拡張が観察されるか。3) TOEIC スコアを基準にエッセイ作成者のグループを三つに分け、その上位と下位グループの間に意味拡張の使用頻度と英語の熟達度が関係するのか。

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

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

【研究発表3】

英語学習者コーパス構築のためのタスク設計：特定の文法項目抽出に向けて

工藤 洋路（玉川大学）

内田 諭（九州大学）

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

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

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

■10月7日（日）

【研究発表第5セッション】

【研究発表1】

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

中谷 安男（法政大学）

1. はじめに

Wood (1982) が主張しているように、国際ジャーナルにおいては、単に結果を報告するだけでなく、その成果が該当分野にどのような貢献をもたらすのか明確な英語で示す必要がある。さらに、研究の価値を明確に読者に報告し、論文の貢献度を訴求する必要がある。そのためには、研究手法に基づき、どのような結果が得られたか正確に伝えなければならない (Cohen, 1994)。これらをまとめて伝えるのは、通常 Discussion と呼ばれる章である。しかし、Swales (2004) が指摘しているように、この箇所を量的に探究した研究は少なく、確固たる証拠よりも、研究者たちの経験や主観でこの章の書き方を示しているものが多い。一つの理由として、この箇所は様々な書き方があり、執筆者の裁量で比較的自由に記述できるという認識がある (例 Swales and Luebs, 2002)。この章では、研究の成果をアピールする必要がある。この際、研究分野の権威者である査読者に主張を伝える際に、自論に関する指摘を予め防御する必要がある。このような時に活用とされるのがヘッジ (Hedge) 表現である。しかし、非母語話者にとって、英語論文の執筆の際に、いかなるヘッジ表現をどのように使用すればよいのか習得するのは容易でない (Charles, 2006)。より具体的なヘッジの考察の章における使用方法を提示すべきである。以上のような観点から本論は、社会科学、人文科学、自然科学分野の代表的な英語による学術論文の比較規模の大きなコーパス・データを分析することにより、Discussion の章に関するこれまでの主張や概念を検証していく (例 Biber et al., 2002)。この際、既存では特定のヘッジ表現の頻度や使用傾向や目的があまり明確ではない。このため、コーパス分析により適確な言語データを抽出し、より説得力のある Discussion の望ましいヘッジ使用形式を提示したい。

2. 検証方法

研究論文の一般的な傾向をみるため、自然科学、社会科学の経済・経営、人文科学の応用言語学から、それぞれインパクトファクターの高い代表的な学術誌を2つずつ選んだ。具体的には、*Science*, *Nature*, *International Economic Review*, *Journal of Management*, *Modern Language Journal*, *Language Learning* の6誌の2006年より2011年に掲載された研究論文の中から、第一著者が英語ネイティブと思われる17本をそれぞれ選んだ。これらを電子ジャーナルからダウンロードしテキストファイルに変

換した。この合計102本の論文による総語数105万語の学術論文コーパスを作成した。

この中の Discussion として明記している章、または明記されていない場合は、それと同等の章の総計79,876語を抜き出した。この章のコーパスを、学術論文コーパスの他の部分を参照コーパスとして、*AntConc Windows 3.5.7*版を活用し特徴語を抽出し、さらにクラスター表現を確認した。また Hyland (2005) の Hedge のリストを活用し、Discussion におけるこれらの表現の頻度と、他の章の参照コーパスにおける頻度と比較した。

3. 結果

アカデミックライティングでは、編集者や査読者を納得させるに、巧みに成果を伝える必要があると考えられている。権威者である読者と文章で交渉を行う英語の記述が必要となる。本研究の成果では、このような目的を達成するために様々なヘッジ表現が使われていることが明らかになった。特に助動詞の *may, might, would* などが、実際のテキストデータでは考察のムーヴを組み立てるのに有効であることが示された。本発表では、これらの分類に該当する事例を検証していく。

【研究発表2】

強意語的機能を持つ罵倒語の進化特性について

新井 洋一 (中央大学)

英語の罵倒語 (expletives or swear words) には、例えば *damn, fuck, bollocks, shit* などのように、それぞれ宗教、性行為、身体器官、排泄物に関係するものが多い。今回の発表では、これらを含む罵倒語の中でも、おもに形容詞や副詞の機能を持つ罵倒語である *bloody, fucking, damn* などを取り上げ、これらの統語的、意味的特性と、ここ最近の際立った傾向について考察したい。なお、以下であげられる用例はすべて BNC からのものである。

これらの罵倒語は、たとえば形容詞として生起する場合、不快な (あるいは否定的) 意味を持つ名詞 (句) と共起し、それらを強調する強意語 (intensifier) 的機能を持つ。

(1) a. They must have been a **bloody nuisance**.

b. ... we'd have a **fucking rain forest**.

c. I knew later I had been a **damn fool**.

また、不快な意味を持たない名詞とも共起するが、その場合は、共起名詞に対して不快な意味合いを付帯する意味的韻律 (semantic prosody) 現象が感知される。

(2) a. That **bloody bird** has annoying me for days.

b. My God, we're all victims in this **fucking city**.

c. Let's get the **damn thing** finished.

一方、これらが副詞として機能する場合は、不快な意味や中立的な意味を持つ形容詞や副詞のみならず、次の用例の下線部で示されるように、むしろ快適な（あるいは肯定的）意味を持つ形容詞や副詞と共起する場合がある。以下では形容詞の例のみあげる。

(3) a. **Bloody** amazing it was!

b. I think he is **fucking** brilliant.

c. And Eleanor was **damn** lucky to have him as an escort once in a blue moon.

また次の用例で示されているように、形容詞のみならず、動詞（句）の強意語として生起している場合もある。

(4) a. I'll **bloody** come down and ask you then.

b. Gary and Margot just **fucking** kill me!

c. That's what they **damn** want, isn't it?

この発表では、(3) や (4) のような例について既存研究を概観し、これらについて深く掘り下げた研究がほとんどない点をまず指摘した上で、BNCとそれ以降の大規模英語コーパスからの用例を鳥瞰し分析を試みたい。その結果として、最近の約30年間のあいだに進行している、罵倒語と共起する [+PLEASANT] な意味素性を持つ形容詞や副詞の種類が増加傾向と、形容詞や副詞のみならず、動詞（句）とさえも共起する副詞機能のいっそうの進化傾向を明らかにし、その文法的仕組みについて考察する予定である。

■10月7日（日）

【研究発表第6セッション】

【研究発表1】

Reliability and Replicability of Annotation Schemes for Learner Corpora

Aika Miura (Tokyo University of Agriculture)

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

In the present study, the author examined learner productions during shopping role plays from the NICT JLE Corpus, which contain 68, 114, and 66 files of utterances at the Common European Framework of Reference for Languages (CEFR) levels of A1, A2, and B1, respectively. In the first scheme, the author identified pragmalinguistic features of requests, drawing on the Cross-Cultural Study of Speech Act Realization Patterns (CCSARP) coding scheme (Blum-Kulka, House, and Kasper, 1989): desire verbs (e.g., *want*) and imperatives were categorized as direct strategy, while ability/permission modals (e.g., *can*) and suggestory (e.g., *how about*) were grouped as conventionally indirect strategies. The second and third schemes were involved with classification of the function of every utterance. The utterances were divided into two major functions: “dealing with transaction” (including subcategories such as “expressing their intention of purchase” and “expressing or asking about the item”) and “communication for transaction” (including “explaining the background,” “requesting an action,” and “confirming”). The degree of grammatical accuracy and discoursal acceptability was also assigned to each utterance: “high” indicated that the utterance was grammatically accurate and acceptable in terms of discourse; “low” was further categorized into (i) coherent (i.e., discoursally coherent but slightly ungrammatical), (ii) slightly incoherent (i.e., semantically inferable but grammatically unacceptable), (iii) incoherent (either discoursally unacceptable due to ungrammatical features or structurally and semantically acceptable but completely incoherent in terms of discourse), and (iv) featuring the use of Japanese. Table 1 shows the total numbers and ratios of the annotated segments.

Table 1

Total numbers and ratios of the annotated segments

The CEFR Level	A1	A2	B3
Total segments for the 1 st scheme	597	1,170	412
Total segments for the 2 nd and 3 rd schemes	893	1,911	1,159
The ratios of the functions			
Dealing with transaction	55.03%	59.2%	0.65%
Communication for transaction	44.97%	40.8%	99.35%
The ratios of the high and low segments			
High	52.18%	54.98%	66.82%
Low			
Coherent	41.72%	41.74%	31.88%
Slightly incoherent	3.68%	2.13%	1.11%
Incoherent	2.18%	1.09%	0.19%
Japanese	0.23%	0.05%	0 %

To attain the reliability and replicability of annotation schemes, the author: (i) documented manuals, (ii) provided annotation training to an external annotator (i.e., checker), (iii) conducted a random check of annotated segments following face-to-face discussions with the checker (i.e., 49.9% and 21.77% of the whole annotations were double-checked by the author and checker, respectively), and (iv) requested the checker to replicate the annotations of 12 files for each annotation scheme without referencing the manuals, followed by some practice. Krippendorff's *alpha*, appropriate for "semantic and pragmatic features" involved with "different magnitudes of disagreement" (Arstein and Poesio, p. 564), were 0.842, 0.862, and 0.481 for the first, second, and third schemes, respectively. Even if it was highly revised after process (iii), the third scheme obtained only the lowest reliability. The discrepancy between the annotators may be attributed to ambiguous definitions in the manuals and the checker's insufficient training, as well as to the overly detailed and complicated classification schemes.

【研究発表2】

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

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

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

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

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

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

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

There seems to be a big difference between the two different learner corpora in terms of what grammatical items count for discriminating the levels and how effective they are. The difference may be partly due to the different tasks in the two corpora, but it should be noted that apparently different modes (i.e. spoken vs. written) require different criterial features.

We will discuss in detail learners’ use of grammatical items as possible candidates of criterial features for their CEFR(-J) levels, together with some methodological and pedagogical implications.

【講演】

Measures of Productivity and Lexical Diversity

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

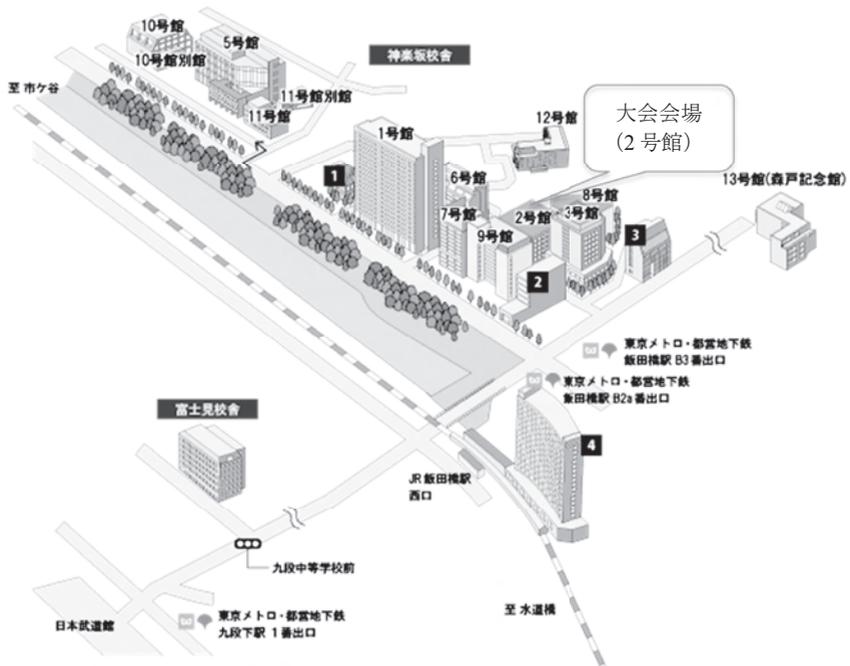
Quantitative measures of productivity and lexical diversity—such as the type-token ratio (TTR), Baayen’s productivity index P or Yule’s K—play an important role in many corpus studies. They have been used to assess the degree of morphological productivity, to estimate the size of an author’s vocabulary,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

a feature in the multivariate analysis of linguistic variation.

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

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

《会場案内図》



《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。
- ・第1日目のワークショップの受付は2号館1階211教室前で9時30分から行います。
- ・大会受付は、第1日（10月6日）は2号館1階211教室前で12時から行います。第2日（10月7日）は9時30分から行います。
- ・構内での喫煙は1号館横の喫煙室にてお願いいたします。
- ・展示について：第1日（10月6日）午後より、221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。
- ・昼食について：第1日（10月6日）は、学生食堂が利用できます。第2日（10月7日）は学内の食堂は営業していませんので、近隣の飲食店をご利用ください。

- ・非会員の参加について：会員ではない方も、「当日会員」としてご参加いただけますので、お誘い合わせの上ご参加下さい（参加費2,000円、2日間共通）。懇親会（下記）へもぜひご参加下さい。大会当日に入会受付もいたします（年会費：一般6,000円、学生3,000円）。
 - ・大会第1日の学術プログラム終了後の懇親会は、インフォーマルな雰囲気の中、参加者同士さまざまな意見交換、情報収集ができる場です。大会ご出席の方々には、ぜひ奮ってご参加いただけましたら幸いです。なお、会場準備の都合上、参加ご希望の方には事前の予約をお願いしております。ご協力のほどよろしくお願い申し上げます。
 - ・英語コーパス学会第44回大会懇親会
 - ・日時：10月6日（土）18:00-20:00
 - ・場所：イタリア食堂 TOKABO 神楽坂店（東京都新宿区神楽坂2-2 志満金ビル4階）
 - ・会費：5,000円
- ※懇親会参加ご希望の方は、参加申込 Web フォーム (<https://goo.gl/forms/cCOJmwiadShD9iN83>) から9月28日(金)までにお申し込み下さい。期限前でも65名の申し込みがあった場合には締め切りますので、お早めにお申し込みください。

『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 本会はコーパスを用いた英語およびその関連領域の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. 大会の開催(年1回)
 2. 会誌・会報の発行
 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、およ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、大学また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. 名誉会員は本会の活動に特別に寄与したものとする。
- 第5条 本会の会員は所定の会費を納めるものとする。会費の額については別にこれを定める。
- 第6条 本会に次の役員をおく。
1. 会長 1名
 2. 副会長 1名
 3. 事務局長 1名
 4. 理事 若干名
 5. 監事 1名
 6. 会計 1名
 7. 顧問 若干名
- 第7条 会長は、理事の互選によって選出する。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 会長は本会を代表し、会務を統括する。会長は総会・理事会を招集し、これを主宰する。

- 第8条 副会長は、理事会の承認を受け会長が委嘱する。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 副会長は会長を補佐し、必要に応じて会長の職務を代行する。
- 第9条 事務局長は、理事会の承認を受け会長が委嘱する。事務局長の任期は2年とし、引き続き2期までの再任を妨げない。
- 2 事務局長は本会の運営に関わる諸事務を担当する。
 - 3 事務局長は会務を円滑に遂行するため、理事会の承認を受け事務局補佐を置くことができる。
- 第10条 理事の委嘱は、理事会の承認を受け会長が行う。理事の任期は2年とし、引き続き再任を妨げない。
- 2 理事は理事会を組織し、本会の運営に関わる重要事項を審議する。
- 第11条 監事は、理事会の承認を受け会長が委嘱する。監事の任期は2年とし、引き続き2期までの再任を妨げない。
- 2 監事は本会の財産および会計の状況を監査する。
- 第12条 会計は、理事会の承認を受け会長が委嘱する。会計の任期は2年とし、引き続き2期までの再任を妨げない。
- 2 会計は本会の会計全般を担当する。
- 第13条 本会に顧問をおくことができる。顧問は役員を退任したもので、本会の発展に特に功績のあったものとする。
- 第14条 本会は毎年1回総会を開く。総会は会則の改定、予算・決算その他重要事項を審議する。
- 第15条 本会に次の委員会を置く。各委員会の規程は別に定める。
1. 編集委員会
 2. 学会賞選考委員会
 3. 大会企画委員会
- 第16条 本会に研究会を置く。研究会の規程は別にこれを定める。
- 第17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。

付則	この会則は2000年4月1日から施行する。
付則	この会則は2009年4月1日から施行する。
付則	この会則は2010年4月1日から施行する。
付則	この会則は2011年4月1日から施行する。
付則	この会則は2014年4月1日から施行する。
付則	この会則は2015年4月1日から施行する。
付則	この会則は2017年4月1日から施行する。
付則	この会則は2018年4月1日から施行する。

細則（第5条）本会の会費は次の通りとする。

1. 一般会員 年額 6,000円（在外会員は年額12,000円）
2. 学生会員 年額 3,000円（在外会員は年額10,000円）
3. 団体会員 年額 5,000円
4. 賛助会員 年額15,000円
5. 名誉会員ならびに顧問からは会費を徴収しない。
6. 会費は年度始めに納入するものとする。また、原則として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。

英語コーパス研究 (第26号)

【2020年2月10日発行】

編集・発行 ©2019 英語コーパス学会
〒157-8511 東京都世田谷区成城6-1-20
成城大学 社会イノベーション学部 石井康毅研究室気付
E-mail (事務局長) : jaecs.hq@gmail.com
Twitter: @JAECS2012
Website: <http://jaecs.com/>
郵便振替口座 : 00930-3-195373 (英語コーパス学会)

印刷所 株式会社ニシキプリント
〒739-2117 東広島市高屋台2-1-12

English Corpus Studies: Vol.26 2019

Articles

Katsumi GOTO / A Study Using Multivariate Analyses of “Authorship” of the Continuation by Thomas Power James to The Mystery of Edwin Drood, and the Evaluation of the Effectiveness of the Word-lists Used in the Analyses.....	1
Shūngo KANAZAWA / The Semantic Differences between Live a/an Adj Life and Lead a/an Adj Life in English	21
Satoru TSUKAMOTO / The Time Span Required for Syntactic Changes in the History of English	39
Conference Program & Abstract	
The 44th Conference of Japan Association for English Corpus Studies	63