



現代英語における前置詞の計量的分析 : 英米差と ジャンル差をふまえた重要前置詞の選定

中西, 淳

(Citation)

統計数理研究所共同研究レポート, 414:107-128

(Issue Date)

2019-03

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90005921>



現代英語における前置詞の計量的分析

—英米差とジャンル差をふまえた重要前置詞の選定—

中西 淳(神戸大学・大学院生)

atsushi.aout11@gmail.com

A Quantitative Analysis of Prepositions in Contemporary English

—Extracting Common Prepositions from Various Contexts—

NAKANISHI Atsushi (Kobe University, Graduate Student)

概要

一般に、名詞や動詞といった内容語に比べて、前置詞の範囲は明確であるとされている。しかしながら、先行研究において、どこまでを前置詞とみなすかについては様々な立場が存在し、また、前置詞の出現状況は英米差やジャンル差のようなテキストタイプによって異なると言及されている。そこで、本研究では、どのようなテキストタイプによって前置詞の出現状況が変化するかを調査した上で、あらゆるテキストタイプで安定して用いられる重要前置詞を特定することを目指す。分析の結果、前置詞の出現頻度はアメリカ英語よりもイギリス英語の方が 1.04 倍高く、話し言葉などのインフォーマルなテキストに比べ、学術論文などのフォーマルなテキストでは 1.43 倍高くなることが確認され、また、英米差よりもジャンル差の方がより本質的な影響を及ぼしている可能性が示された。最後に、これらの結果をふまえて、これらすべてのテキストタイプの頻度を合成し、安定して高頻度で使用する「重要前置詞」を抽出した。

キーワード

前置詞, 英米差, ジャンル差, コーパス

1. はじめに

日本人英語学習者は前置詞をどの程度使用できているであろうか。この問いは一見平易なように見えて、実際には計量的に議論することがきわめて難しい。というのも、先行研究において、前置詞の定義は一定しておらず、前置詞全体の頻度を計測することはほとんど不可能であると言える。この点をふまえると、前置詞研究を計量的な観点から行っていく場合、まずもって重要なことは、現代英語の様々なテキストタイプにおいて、安定して高頻度で使用する前置詞を特定し、自らの議論の対象を明確に定義することである。このことは、前置詞研究を科学的に進めていくための議論の基盤を構築する上で不可欠であると言える。

もっとも、名詞や動詞に比べると、前置詞はどのようなテキストタイプにおいても安定して高頻度

で出現するようになる。しかしながら、実際には、前置詞の出現状況はテキストタイプによって一定の範囲で変化する。以下は、その一例である(下線は筆者、以下同)。

- (1) a. I heard someone coming towards the door. (BNC_FICTION)
- b. I heard her coming toward the door, ... (COCA_FICTION)
- c. ..it's based on information that's happening out there today.. (COCA_SPOKEN)
- d. Based upon this information, it is indicated that as teacher educators we have begun to professionalize... (COCA_ACADEMIC)

上記は、イギリス英語を収集した British National Corpus (以下, BNC)とアメリカ英語を収集した Corpus of Contemporary American English(以下, COCA)の小説セクション(FICTION), 話し言葉セクション(SPOKEN), 学術論文等セクション(ACADEMIC)から抜粋した用例である。まず, (1a)と(1b)に注目すると, イギリス英語では *towards* が, アメリカ英語では *toward* が使用されている。また, (1c)と(1d)に注目すると, 話し言葉では *on* が, 学術論文等では *upon* が使用されている。下線部の語は異なる前置詞であるにも関わらず, これらの語の間に意味の差は存在しない。事実, 学習用英英辞典の代表格である *Longman Dictionary of Contemporary English* (LDOCE)を見ても, これらの語の意味の違いに明確な言及はなく, 「イギリス英語では *towards* が, アメリカ英語では *toward* が好まれる」, 「一般的には *on* を使用するがフォーマルな場面では *upon* が好まれる」という指摘が存在する。

以上をふまえると, 英米差やジャンル差が前置詞の出現頻度に一定の影響を及ぼす蓋然性が高い。しかしながら, この点をふまえ, 多様なテキストタイプにおける前置詞頻度を調査し, それらを総合することで, 教育や研究において重要な前置詞を特定しようとした試みは管見の限り存在しない。そこで, 本研究は, 英米差やジャンル差をふまえた前置詞の頻度調査を行い, それらを統合することで以後の議論の対象となる「重要前置詞」の特定を目指そうとするものである。

2. 先行研究

前置詞の用法に関する研究は多くなされているが, テキストタイプによって前置詞の使用状況や, その機能がどのように異なるかを計量的に調査したものは筆者の知る限り多くない。ここでは, まず, 前置詞使用の英米差やジャンル差に関する研究のうち, 代表的なものを概観する。

1 点目は, Francis & Kučera (1982)がある。この研究では, 彼ら自身が構築したアメリカ英語コーパスである Brown Corpus と, 同様のデータ収集方法によって構築したイギリス英語コーパスである LOB Corpus の 2 つを用い, 前置詞を含む 9 種の品詞の使用割合を比較し, 以下の結果を得た。

表 1 主要品詞の使用割合(英米差)

	Brown Corpus (%)	LOB Corpus (%)
名詞	26.8	25.2
動詞	18.2	17.8
限定詞・数量詞	14.2	14.2
前置詞	12.0	12.2
形容詞	7.1	7.3
代名詞	6.6	7.1
接続詞	5.9	5.5
副詞	5.2	5.5
その他	4.0	5.2

上表からも確認されるように、全品詞における前置詞の割合は、アメリカ英語では 12.0%であるのに対し、イギリス英語では 12.2%であり、イギリス英語で前置詞がやや多用される傾向が見られた。しかしながら、この研究は前置詞のみに焦点を当てた研究ではないため、英米間で前置詞の使用頻度が異なる原因については言及されていない。

2 点目は、Kennedy (1998)である。この研究では、Francis & Kučera (1982)と同様に、Brown Corpus と LOB Corpus を用いて、アメリカ英語とイギリス英語において、高頻度上位 14 語の前置詞が、前置詞全体頻度のうち、どの程度の割合を占めているかを調査した。結果は以下の通りであった。

表 2 前置詞全体頻度に占める主要前置詞の使用割合(英米差)

	Brown Corpus (%)	LOB Corpus (%)
of	29.8	28.6
in	17.1	16.4
to	9.1	8.8
for	7.4	7.1
with	6.0	5.8
on	5.1	5.1
at	4.3	4.4
by	4.3	4.6
from	3.6	3.8
into	1.5	1.3
about	1.0	1.0
through	0.7	0.6
over	0.7	0.6

between	0.6	0.7
total	91.3	88.8

まず、前置詞全体頻度における 14 種前置詞の総出現割合に関しては、アメリカ英語では 91.3%、イギリス英語では 88.8%であった。その差は大きなものではないが、イギリス英語の方が多様な前置詞を好むのに対し、アメリカ英語の方が重要前置詞をより集中的に使用している傾向が示唆された。また、個別前置詞については、on や about のように英米どちらにおいても頻度比率が変わらないものや、through や over のようにアメリカ英語でより頻度比率の高いもの、by や between のようにイギリス英語でより頻度比率が高いもの、などが存在することが確認された。一般に、英米差があまりないと考えられる重要前置詞に関して、英米差が明確に存在することを指摘した点で、この研究は重要なものであると言える。

3 点目は、Mindt & Weber (1989)である。この研究では、様々な前置詞のうちイギリス英語・アメリカ英語の両方で用いられるもの、イギリス英語のみで使用されるもの、アメリカ英語のみで使用されるものをそれぞれ抽出した。結果は以下の通りであった。

表 3 イギリス英語・アメリカ英語のどちらにも含まれる前置詞／どちらかのみに含まれる前置詞

イギリス英語・アメリカ英語どちらも	イギリス英語のみ	アメリカ英語のみ
of, in, to, for, with, on, at, by, from, into, about, through, over, between, after, under, against, during, without, than, out, upon, before, toward, per, among, within, around, across, behind, along, down, above, since, upon, except, including, beyond, off, near, throughout, but, until, as, despite, outside, inside, below, beside, towards, concerning, onto, past, beneath, versus, regarding, like, besides, considering, opposite, plus, via, unto, amid, following, alongside, underneath, aboard, save, pending, amongst, minus, round, till, amidst, nearest, notwithstanding, times, nearer, respecting, excepting	worth, barring, less, failing, re, touching, bar, afore, à la, bout, qua, vice, neath	unlike, pursuant, excluding, pro, astride, atop, involving, post, vis-à-vis, dell, infra, inter, with-but-after

上表に見られるように、イギリス英語・アメリカ英語両方で使用される前置詞は 81 種存在するのに対し、いずれか一方のみで使われる前置詞は 13 種ずつであった。多くの前置詞は英米差なく使用できるが、一部の前置詞は明確にいずれかの言語変種に強く結びついている可能性があると言える。

以上の 3 つの研究で見たように、英米差は前置詞の頻度に影響を及ぼす要因の 1 つであると言える。しかしながら、前置詞の頻度は内容ジャンルによっても影響されることが考えられる。以下では、ジャンル差が前置詞頻度に対して及ぼす影響について調査した 2 つの主要研究について概観する。

1 点目は、前述の Francis & Kučera (1982) である。彼らは、英米差だけではなくジャンル差についても調査を行っており、Brown Corpus と LOB Corpus に含まれるテキストを情報系テキストと小説系テキストの 2 種類に分類し、9 種の品詞の出現割合を比較した。

表 4 主要品詞の使用割合(ジャンル差)

	情報系テキスト (%)	小説系テキスト (%)
名詞	27.7	20.9
動詞	16.7	21.8
限定詞・数量詞	15.0	11.8
前置詞	12.9	9.7
形容詞	7.7	5.5
代名詞	4.9	12.5
接続詞	5.7	5.6
副詞	4.9	7.0
その他	4.5	7.1

上表からも確認されるように、情報系テキストのうちの 12.9% が前置詞であるのに対し、小説系テキストでは 9.7% であり、事実を伝えるノンフィクションのテキストにおいて前置詞が多用されることを明らかにしている。

2 点目は、Leech, Rayson & Wilson (2001) である。この研究では、冒頭で触れた BNC を用い、話し言葉と書き言葉、情報系テキストと小説系テキストで特徴的に使用される品詞を特定した。結果は以下のようになった。

表 5 各テキストにおいて特徴的に見られる品詞タグ上位 4 種

話し言葉	書き言葉	情報系テキスト	小説系テキスト
UH (感動詞)	NP1 (単数固有名詞)	NN2 (複数形名詞)	PPIS1 (一人称単数代名詞)
FU (分類不可)	NN1 (単数名詞)	MC (数量詞)	PPY (二人称代名詞)
PPIS1 (一人称単数代名詞)	JJ (形容詞)	JJ (形容詞)	VVD (動詞の過去形)

PPY (二人称代名詞)	NN2 (複数名詞)	IO (前置詞 of)	PPHS1 (三人称単数代名詞)
-----------------	---------------	----------------	---------------------

上表に示されるように、前置詞は話し言葉と書き言葉の区別の観点とはなっていないが、情報系と小説系の対比においては、情報系テキストを特徴付ける品詞に前置詞 of が含まれている。

最後に、いわゆる重要語抽出に関わる研究について概観しておく。前置詞に限って重要語リストを開発しようとした研究は知る限りほとんど存在しない。しかしながら、たとえば、大学英語教育学会(JACET)が大学生向けに公表している語彙表「JACET 8000」では、重要語の中にいくつかの前置詞が含まれている。以下は、2003 年に出版された第 3 版と、2016 年に出版された第 4 版における前置詞の頻度順である。

表 6 JACET8000 における前置詞トップ 10

順位	第 3 版	順位	第 4 版
3	to	2	of
4	of	5	in
6	in	6	to
11	for	11	for
16	on	14	with
19	with	15	on
26	at	16	at
27	from	25	from
34	by	26	by
38	about	46	about

第 3 版は、BNC のみを利用し、BNC 頻度と日本人英語学習者が遭遇しやすい英語資料の両方で頻度をとり、頻度がずれている場合は日本人用の英語資料における頻度を加味する形で重要語リストを作成している。一方、第 4 版では、BNC のみならず、COCA も利用し、かつ、それぞれのジャンルごとの頻度を取り出して、頻度を合成した上で、日本人向けの資料と比較して順位を決定している。上表からも確認されるように、いずれの版においてもこれら 10 種の前置詞は 50 位以内にすべて含まれており、英語全体の中でも前置詞は重要語としてみなされていると考えられる。このように、コーパスに準拠した重要語リストの作成は英語教育において広く行われており、その中でも前置詞は重要な位置づけであることが確認される。

以上の研究で見たように、先行研究では英米差やジャンル差が前置詞使用に影響を及ぼす可能性を明らかにしてきた。これらの知見は有益なものであるが、いくつかの課題も残されている。ここでは、特に重要であると思われる 4 点について指摘する。

1 点目は、前置詞として具体的に分析される語にずれが存在することである。たとえば、すでに

述べたように、Kennedy (1998)では、14 種の重要前置詞に限って出現頻度を調査したわけであるが、なぜ 14 種が選ばれたのか、14 種の前置詞以外の出現頻度がどのようになっているかについては明らかにされていない。前置詞の定義は幅広く、その詳細な定義は研究者によって異なる。この点をふまえると、恣意的に選ばれた語群のみを対象とするのではなく、何らかの外的基準に基づいて前置詞を明確に定義した上で、そこに含まれるものを悉皆的に調査していくことが必要になるだろう。

2 点目は、内容ジャンルに関する研究が、きわめて大まかな分類にとどまっているということである。Francis & Kučera (1982)や Leech, Rayson & Wilson (2001)では、あらゆるテキストを小説系と情報系の 2 種に分けて議論を行っている。しかし、ジャンルの分類にはこれ以外にも様々なものが存在し、そうした細かい内容差が前置詞使用にどのように影響を及ぼしているかについては必ずしも明らかにされていない。

3 点目は、英米差とジャンル差のどちらがより前置詞使用に影響しているかが明らかにされていないことである。英米差やジャンル差に関して、別々に調査したものは存在するものの、2 つを同じ範疇で比較し、英米差とジャンル差のどちらがより大きな影響を与えているかについて統計的に説明しようとする研究はあまりない。しかしながら、テキストタイプが前置詞使用に及ぼす影響を総合的に考察する場合、この点の説明がきわめて重要になると言える。

4 点目は、前置詞のみに注目した重要語リストの開発はほとんどされていないことである。英語教育において、学習上の重要語リストを開発する試みは多くなされており、現代英語コーパスを用いた計量的な手法での抽出も広く行われている。一方、前置詞については、いくつかの先行研究で質的に「重要前置詞」を取り上げているものの、その基準は定かではなく、統計的指標を用いた重要前置詞リストを作成することが不可欠であると言える。

3. リサーチデザインと手法

3.1 研究目的と RQ

既に述べたように、本研究では、先行研究において前置詞の頻度に影響を及ぼすと考えられている英米差とジャンル差の 2 点について、その各々が実際に影響を及ぼしているかどうか、また、それらの影響をあまり受けず、あらゆるテキストにおいて高頻度に出現する前置詞があるかどうか、を頻度情報に基づいて統計的に明らかにし、前置詞研究を行う際の議論の基盤を構築することを目指す。これらの点をふまえて、以下のリサーチクエスチョン(RQ)を設定する。

RQ1 英米差は前置詞の頻度に影響しているか？

RQ2 ジャンル差は前置詞の頻度に影響しているか？

RQ3 英米差とジャンル差ではどちらがより前置詞の頻度に大きく影響しているか？

RQ4 英米差やジャンル差を越えて安定して高頻度に出現する重要前置詞は何か？

3.2 使用するデータ

本研究では、すでに述べた BNC と COCA を用いることとし、ブリティッシュ大学(BYU)の Mark Davies 氏が運営する BYU-BNC と BYU-COCA の 2 つのサイトを経由してデータにアクセスを行う。

これらのコーパスは、いずれかの分野に偏ることなく、幅広いテキストタイプからデータを収集しており、英語コーパスの中でも学術的に信頼度の高いものであるとされているが、これら 2 つのコーパスを比較するにあたって留意すべきことが 3 点ある。

まず、1 点目は、ジャンルの違いである。これら 2 つのコーパスはそれぞれジャンル別に区分されているが、その区分は、BNC が SPOKEN(話し言葉), FICTION(小説), MAGAZINE(雑誌), NEWSPAPER(新聞), NON-ACADEMIC(非学術), ACADEMIC(学術), MISCELLANEOUS(雑多)の 7 種類であるのに対し、COCA は SPOKEN, FICTION, MAGAZINE, NEWSPAPER, ACADEMIC の 5 種類である。そこで、本研究では、比較対象をそろえるため、両コーパスに共通する SPOKEN, FICTION, MAGAZINE, NEWSPAPER, ACADEMIC の 5 つのジャンルのデータのみを調査対象とした。以下、分析に使用するデータの頻度情報を 10 種(英米 2 種×ジャンル 5 種)に分けて示す。

表 7 各ジャンルの総語数

	BNC	COCA
SPOKEN	9,963,663	116,748,578
FICTION	15,909,312	111,845,122
MAGAZINE	7,261,990	117,354,113
NEWSPAPER	10,466,422	112,995,407
ACADEMIC	15,331,668	111,410,528

2 点目は、上記で選定したデータの総量が異なるという点である。BNC では 5 つのジャンルを合わせて約 6 千万語であるのに対し、COCA は約 6 億語であり、2 つのコーパスの総語数には 10 倍程度の差が存在する。そこで、本研究では、2 つのコーパスの頻度を比較するにあたって、粗頻度ではなく調整頻度を使用することとした。調整頻度には様々な単位が存在するが、ここでは 100 万語当たりの調整語数(PMW)を使用する。

3 点目は、それぞれのジャンルについて具体的なデータ収集の方法が異なっていることである。たとえば、話し言葉の場合、BNC においては、自然会話を集めた一般話者データと社会生活における場面別データの 2 種類が存在するが、COCA においては、放送番組のスクリプトを中心とするデータで構成されている。そのため、同じ SPOKEN というジャンルに分類されているものでも BNC と COCA のデータは厳密に言えば同じではない。また、こうした違いはすべてのジャンルに関して見られる。しかしながら、この点の対応については今後の課題とし、本研究では、それぞれのコーパスのジャンルを相互比較可能なものであるとみなした上で、以下の分析を行う。

3.3 対象前置詞

すでに述べたように、先行研究の一部は、いわゆる「重要前置詞」を取り上げて研究を行っているが、何をもって重要前置詞とするのかを合理的に定めることは容易ではない。そこで、本研究では、Francis & Kučera (1982)や Leech, Rayson & Wilson (2001)のアプローチにならい、品詞タガーによって前置詞と判定されたものすべてを対象とすることとした。なお、ここで問題となるのは、どのようなタガーを使用するかということである。前述の Francis & Kučera (1982)では Penn Treebank を、Leech, Rayson & Wilson (2001)では CLAWS (the Constituent Likelihood Automatic Word-tagging System)を使用していたわけであるが、本研究では、BYU のシステムで採用されている CLAWS を使用する。なお、このシステムにおいて、いわゆる前置詞には「_i*」のタグが付与され、このタグが付与された語は全体で 3,000 種以上あった。

BYU のシステムで採用されているタガーは一般に 96~97%の精度を持つとされているが、実際には、品詞の判定ミスも存在する。たとえば、以下は、「_i*」のタグが付いたもののうち頻度が 1 回の語の事例である。

- (2) a. At 0749 Fuchida sent the signal' To-to-to'...(BNC_FICTION)
b. On the front panel of the selected case can be mounted the main On-Off switch...
(BNC_FICTION)

上記の(2a)は信号音を表すオノマトペである“To-to-to”が、(2b)はスイッチの機能を示した語である“On-Off”がそれぞれ独立した前置詞と誤解析されており、これらは明らかに前置詞とは言えない。このように、前置詞のタグが付与されていても頻度のきわめて低い語は前置詞となっていない可能性が高い。そこで、本研究では、便宜的に頻度 5 という閾値を設定し、頻度 5 を下回る語をすべて分析対象から除外した。これにより、当初の 3,000 種を上回る前置詞の候補語は 220 種まで整理された。

次に、これら 220 種の語を目視で確認し、不適切なものが混入していないかを調べた。その結果、品詞解析のミスであると思われるものが 6 種、一般的な辞書の記述に含まれていないものが 5 種確認された。(3a)は前者、(3b)は後者の例である。

- (3) a. Inland-Nippon's I-N Tek, the world's most modern cold-rolling steel mill...
(COCA_NEWSPAPER)
b. Today I had a quiet restful lunch chez Ira Mazer...(COCA_FICTION)

(3a)は固有名詞が、(3b)は at を意味するフランス語がそれぞれ前置詞として扱われており、これらを前置詞とみなすことは適切でない。そこで、本研究では、これら 11 種をすべて分析対象から除外した。これにより分析対象となる前置詞は 209 種となった。これらは計量的な観点から前置詞を

議論する際の最大の範囲であるが、その中には限られたテキストタイプのみでしか使用されないものも含まれている。つまり、前置詞の議論に妥当性を持たせるためには、この 209 種の中からあらゆるテキストタイプにおいてある程度共通して出現する語を絞っていくことがきわめて重要であると言える。

また、これらの前置詞にはテキストタイプ間での頻度の違いのみならず、語形成上の違いも存在する。本研究では、小西(1976)の枠組みを参考に、209 種の前置詞を、「一次前置詞」、「複合前置詞」、「他品詞由来前置詞」、「群前置詞」の 4 区分に分類した。なお、「群前置詞」は、BYU システムにおいて、かたまりで 1 語と数えられておらず、個別にタグが付与されている(例: “in_front_of” は “in_i*”・“front_i*”・“of_i*” の 3 語に前置詞タグ)。そこで、本研究では、個別前置詞(in や on)は当該前置詞の頻度を除算して分析を行った。以下、それぞれの分類区分ごとの種別数 (type) と総語数 (token) を例とともに示す。

表 8 前置詞分類

小西 (1976)	type	token (%)	例
一次前置詞	26	82.8%	of, in, to
複合前置詞	33	6.8%	about, into, half-past
他品詞由来前置詞	62	8.5%	near, during, per
群前置詞	88	1.7%	in_front_of, next_to
計	209	100%	

3.4 手法

まず、RQ1 では、イギリス英語とアメリカ英語を、(1)総頻度、(2)4 区分の占有比、(3)高頻度前置詞上位 10 語、の 3 つの観点で比較する。このうち、(1)総頻度については、前置詞頻度がイギリス英語とアメリカ英語で異なっているかを調査するため、BNC と COCA に出現する前置詞総頻度を比較する。また、差が有意であるかどうかを検証するために、カイ自乗検定を行う。検定には統計解析言語である「R version 3.5.1」を使用し、あらかじめインストールされている `chisq.test` 関数を用いて分析を行う。2 つのデータセットにおいて、頻度に差がないという帰無仮説を否定する危険率(p 値)に注目することは一定の妥当性を持つが、最近の研究では、 p 値がデータサイズに直接的に依存することから、 p 値のみで議論を進めていくことに批判的な立場も存在する(竹内・水本, 2012)。そこで、本研究では、危険率にくわえ、その差が実際にどの程度具体的な意味を持つのかを示す効果量にも注目することとした。小林 (2015)において、頻度差検定を行う際の効果量について、 ϕ 係数(phi coefficient)やクラメールの V (Cramér's V)、オッズ比(Odds Ratio: OR)の 3 つが紹介されているが、本研究では、オッズ比を用い、その 95%信頼区間(95% Confidence interval: CI)の値が 1 をまたいでいないかどうか判断の基準にする。なお、オッズ比の計算は R の `vcd` パッケージに含まれる `oddsratio` 関数を用いる。

つづいて、(2)4 区分の占有比については、具体的に、どのような種類の前置詞がイギリス英語と

アメリカ英語でより多く用いられているかを特定するため、小西(1976)の「一次前置詞」、「複合前置詞」、「他品詞由来前置詞」、「群前置詞」の4区分を用いる。それぞれの区分ごとの前置詞の出現頻度を合算し、その後、それぞれの区分ごとの占有比を算出し、それらをイギリス英語とアメリカ英語で比較する。

最後に、(3)高頻度前置詞上位10語については、高頻度に用いられている前置詞が英米間での程度一致しているかを調査する。そのため、イギリス英語とアメリカ英語で出現頻度のもっとも高い前置詞上位10語を抽出し、その一致度を見る。

また、RQ2においても、RQ1と同様、(1)総頻度、(2)4区分の占有比、(3)高頻度前置詞上位10語、の3つの観点で比較する。このとき、比較するデータとして、ジャンルごとにBNC頻度とCOCA頻度を合算した値を求める。(1)総頻度については、RQ1と同じ手順で進めていくが、5つのジャンル間での比較を行うため、まず、クロス集計表に対してカイ自乗検定を行い、ジャンル間のどこかに頻度の差があるかどうかを確認する。その後、差があると確認された場合、ボンフェローニ法を用いた下位検定を行い、個々のジャンル間に差があるかどうかを確認する。(2)4区分の占有比についても、RQ1と同様、4区分ごとの前置詞の出現頻度を合算し、その後、それぞれの区分ごとの占有比を算出し、それらを5つのジャンル間で比較する。(3)高頻度前置詞上位10語についても、RQ1と同様、5つのジャンルごとに出現頻度のもっとも高い前置詞上位10語を抽出し、その一致度を見る。

次に、RQ3では、英米差とジャンル差のどちらがより前置詞の出現に影響しているか、また、どのような前置詞がそれぞれのテキストにおいて特徴的に使用されているか、を調査する。分析には、多変量解析の1つであるコレスポンデンス分析を行う。第1アイテムはテキストタイプとし、ここには2種の地域差と5種のジャンル差を掛け合わせた合計10種の категорияが存在する。また、第2アイテムは前置詞とし、ここには209種の個別前置詞の категорияが存在する。なお、コレスポンデンス分析においては、RのMASSパッケージに含まれるcorresp関数を用いて分析を行う。

最後に、RQ4では、テキストタイプの影響をあまり受けず、共通して高頻度で出現する前置詞を抽出する。このため、10種のテキストタイプを変数、209種の前置詞をケースとする頻度表に対して主成分分析を実行し、第1主成分得点の高い語を抽出する。なお、分析にあたってはRのprcomp関数を用いる。

4. 結果と考察

4.1 RQ1 英米差

はじめに、(1)総頻度に注目する。イギリス英語とアメリカ英語の2つの地域における前置詞の出現頻度は以下ようになった。

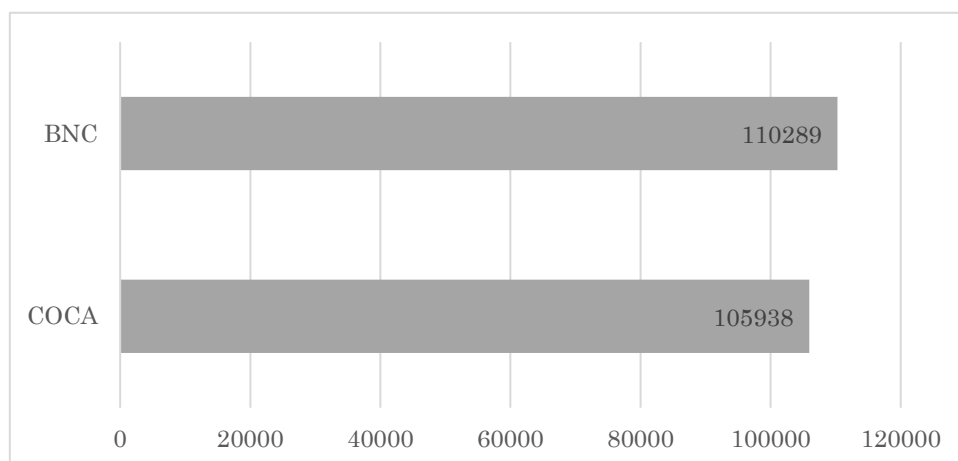


図 1 前置詞総頻度の英米差

上図から確認されるように、BNC が COCA をわずかながら上回っており、イギリス英語の方がアメリカ英語よりも前置詞を多く用いることが確認された。カイ自乗検定を行ったところ、その差は有意であったが、効果量はきわめて小さく、95%信頼性区間は 1 にまたがっていることが確認された($\chi^2=8569.9$, $df=1$, $p<.001$, 95% $CI=0.95-1.13$, $OR=1.04$)。この結果から、先行研究が示す「イギリス英語>アメリカ英語」という前置詞使用量の関係性は確認されたものの、その差はあまり大きなものではないことが明らかになった。

つづいて、(2)4 区分の占有比を英米間で比較する。得られた結果は以下のようになった。

表 9 前置詞種類別の占有比(英米間)

	BNC	COCA
一次前置詞	83.5%	82.9%
複合前置詞	6.6%	7.3%
他品詞由来	8.0%	7.9%
群前置詞	2.0%	1.9%

比率値の差に注目すれば、BNC では、一次前置詞が 0.6%、他品詞由来が 0.1%、群前置詞も 0.1%大きくなっていた。その一方で、COCA では、複合前置詞のみ 0.7%大きくなっていた。この点をふまえると、前置詞全体で見たとき頻度差はあまり大きなものではないことが明らかになったが、注目すべき違いとして複合前置詞の頻度の差が挙げられる。アメリカ英語において複合前置詞が多用される理由としては、*atop* や *onto* のように、“at the top of”や“on to”が複合された前置詞が特徴的に使用されており、アメリカ英語において省略表現が多用されているためであると考えられる。

この違いについて、*atop* と *onto* を例に、省略表現の使用率を BNC と COCA で調査した。その結果、*atop* 率(*atop* 頻度/*at_the_top_of* 頻度)は、イギリス英語では 8.8%であるのに対し、ア

アメリカ英語では 95.8%であり, onto 率(onto 頻度/on_to 頻度)は, イギリス英語で 37.3%, アメリカ英語で 149.1%となっており, アメリカ英語では省略表現が好んで使用される傾向にあり, このことが複合前置詞の多用につながったと考えられる。

次に, (3)高頻度前置詞上位 10 語を英米間で比較する。以下は, その結果である。

表 10 英米別高頻度前置詞上位 10 語

頻度順	BNC	COCA
1	of	of
2	in	in
3	to	to
4	for	for
5	with	with
6	on	on
7	at	at
8	by	from
9	from	by
10	as	about

注目すべきは, 頻度順のいくつかは異なっているものの, 上位 10 語中 9 語が同じ前置詞になっていることである。前置詞の全体頻度を調査したときと同様, 高頻度前置詞に限れば, 個別前置詞の使用に関しても英米間に大きな差異は見られず, 特に, 上位 7 位までの前置詞においては頻度順もまったく同じになっており, これらの前置詞は英米関係なく安定的に使用される前置詞であることが示唆された。

4.2 RQ2 ジャンル差

RQ1 では前置詞使用の英米差について見てきたわけであるが, RQ2 では前置詞使用がジャンルによって異なっているかについて見ていく。

まず, (1)総頻度を 5 つのジャンル間で比較する。得られた結果は以下のようになった。

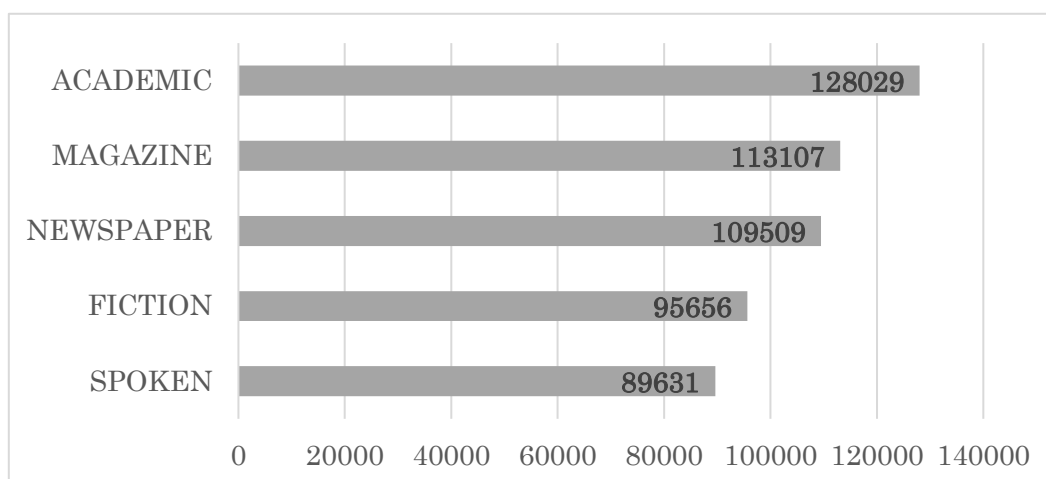


図 2 前置詞全体頻度のジャンル差

上図より、前置詞総頻度について、ACADEMIC > MAGAZINE > NEWSPAPER > FICTION > SPOKEN という順序性が見られた。カイ自乗検定を行ったところ、ジャンル間に差があるということ ($\chi^2=946130$, $df=4$, $p<.001$), また、多重比較により、MAGAZINE と NEWSPAPER を除く、すべてのジャンル間の差が有意であることがそれぞれ確認された (ACADEMIC・MAGAZINE 間: $\chi^2=168690$, $p<.001$, 95% CI=1.17-1.18, OR=1.17, MAGAZINE・NEWSPAPER 間: $\chi^2=82$, $p<.001$, 95% CI=0.99-1.01, OR=1.00, NEWSPAPER・FICTION 間: $\chi^2=107120$, $p<.001$, 95% CI=1.14-1.14, OR=1.14, FICTION・SPOKEN 間: $\chi^2=22814$, $p<.001$, 95% CI=1.06-1.07, OR=1.06)。

以上より、ACADEMIC や、MAGAZINE, NEWSPAPER などの情報系テキストでは前置詞が多用され、SPOKEN や、FICTION などの小説系テキスト、および、話し言葉では前置詞があまり使用されないことが確認された。この結果は、Francis & Kučera (1982) や、Leech, Rayson & Wilson (2001) を支持するものとなった。

では、ACADEMIC や MAGAZINE ではなぜ前置詞の出現頻度が高いのであろうか。以下は、コーパスから得られた用例の一部である。

(4) The level of unemployment in one of Easton's wards in 1981 was 2.5 times higher than that in another... (BNC_ACADEMIC)

このように、ACADEMIC などの情報系テキストでは、1 つの文の中に様々な前置詞が使用され、文要素が多重的に結合している。これに対して、SPOKEN や FICTION のような非情報系テキストでは、接合関係が希薄になっていることから、前置詞があまり使われていないと考えられる。

つづいて、(2)4 区分の占有比をジャンル間で比較したところ、以下の結果が得られた。

表 11 前置詞種類別の占有比(ジャンル間)

	SPOKEN	FICTION	NEWSPAPER	MAGAZINE	ACADEMIC
一次前置詞	82.7%	81.0%	83.5%	83.5%	85.3%
複合前置詞	8.2%	10.5%	6.2%	6.8%	5.4%
他品詞由来	7.5%	7.1%	8.2%	7.8%	6.9%
群前置詞	1.6%	1.5%	2.0%	1.9%	2.4%

全体的に見れば、一次前置詞が 8 割以上を占めている点や、群前置詞の割合がもっとも低い点はなど、5 つのジャンル間で共通した傾向が見られる。一方で、複合前置詞と他品詞由来の順位関係については、SPOKEN と FICTION では複合前置詞の方が多く、NEWSPAPER, MAGAZINE, ACADEMIC では他品詞由来の方が多くなっており、多少の異なる傾向があることが示唆された。

以上、前置詞の全体頻度におけるジャンル差と、前置詞のタイプ別のジャンル差の 2 つについて調査を行ったわけであるが、これらの 2 つの結果とも SPOKEN や FICTION のようなインフォーマルなテキストと MAGAZINE や ACADEMIC のようなフォーマルなテキストで明確な差異が確認され、これらは Francis & Kučera (1982)や、Leech, Rayson & Wilson (2001)の結果を支持するものとなった。

最後に、(3)高頻度前置詞上位 10 語をジャンル間で比較した。以下、その結果を示す。

表 12 ジャンル別高頻度前置詞上位 10 語

	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1	of	of	of	of	of
2	in	in	in	in	in
3	to	to	to	to	to
4	for	with	for	for	for
5	on	for	with	with	with
6	with	on	on	on	on
7	at	at	at	at	by
8	about	from	from	from	from
9	from	into	by	by	as
10	by	by	as	as	at

まず、高頻度上位 3 語は 5 つのジャンルすべて一致していることが確認された。RQ1 で英米それぞれの高頻度前置詞上位 3 語とも一致していることから、これら 3 語は非常に安定的に高頻度で使用される前置詞であると考えられる。

一方、特定のジャンルにのみ高頻度で出現している前置詞も存在する。たとえば、about は

FICTION, MAGAZINE, NEWSPAPER, ACADEMIC では上位 10 語に含まれていないが、SPOKEN でのみ 8 位に位置付けられている。では、なぜ SPOKEN において **about** が多用されるのであろうか。ここでは、以下の 3 つの用例に基づいてその理由を考察する。

- (5) a. That's the only two good points about them. (BNC_SPOKEN)
b. You're talking about(for) half hour there and back drive. (BNC_SPOKEN)
c. Well I think that's about all. (BNC_SPOKEN)

1 つ目は、(5a)に見られるように、「～に関する」という表現において **about** が多用されていることが挙げられる。他のジャンルにおいては、**of** や **on**, **concerning** といった様々な表現が使用されている一方で、話し言葉においては、**about** が多く用いられていると考えられる。

2 つ目は、他の前置詞の代わりに使用されていることが挙げられる。一般に、話し言葉においては前置詞を含む機能語にあまり焦点が当てられていないと考えられ、**about** が他の前置詞の役割を代用することで言語産出の流暢性を高めていると考えられる。たとえば、(5b)において、**about** は **for** の代わりとして使用されており、期間を表す前置詞として機能している。

3 つ目は、(5c)のように、程度を弱める副詞的な用法で使用されていることが挙げられる。**about** は副詞的に使用されることも多く、話し言葉においては、程度を緩和させる働きや、強勢を置く働きとして多く使用されていると考えられる。

なお、これらの知見については、あくまで推測であり、具体的な原因についてさらなる調査を行う必要があると言える。

4.3 RQ3 英米差とジャンル差の関係

RQ3 では、RQ1 と RQ2 の結果をふまえ、英米差とジャンル差を同時に分析することで、前置詞使用に影響があると考えられるこれら 2 つの要因のうち、どちらがより影響しているかを調査する。また、同時に、それぞれのテキストタイプにおいて特徴的に使用される前置詞を特定する。以下、10 種のテキストタイプと 209 種の前置詞の関係性を視覚化したコレスポンデンス分析の結果を示す。

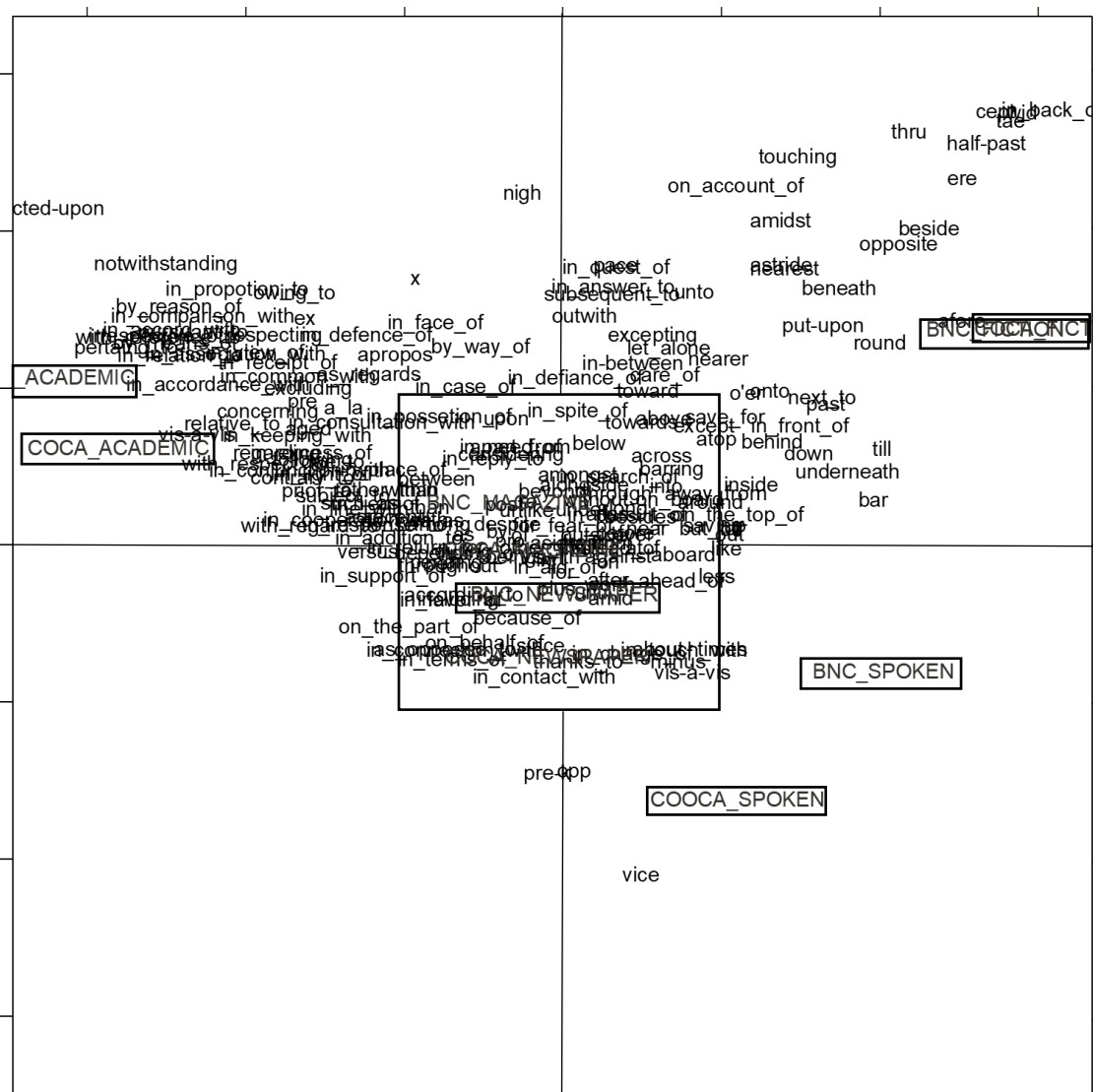


図 3 コレスポンデンス分析の結果(テキストタイプ 10 種×前置詞 209 種)

コレスポンデンス分析の結果、一次元の寄与率は65.17%、二次元の寄与率が15.97%であり、これら2つの次元で80%以上が説明されることが確認された。

上記で得られた2つの次元はさらに、第一象限(右上)、第二象限(左上)、第三象限(右下)、第四象限(左下)の4つの象限に区分される。さらに、散布図の解釈では、原点付近ははっきりとした特徴を持たない領域であるとされる。このため、原点を中心に±1の範囲に入るものをコア域と仮定すると、全体のデータは4つの象限とコア域に区分されることとなる。

では、それぞれの区分にどのようなデータが含まれているのであろうか。以下は、それぞれの区分に布置されたテキストタイプと特徴前置詞の一覧である。

表 13 各象限に付与されたテキストタイプと特徴前置詞

区分	テキストタイプ	特徴前置詞
第一象限	BNC_FICTION COCA_FICTION	位置・方向を示す前置詞(beneath, beside) 省略前置詞(thru, cept)
第二象限	BNC_ACADEMIC BNC_ACADEMIC	群前置詞(in_comparison_with) 動詞由来前置詞(regarding, concerning)
第三象限	×	
第四象限	BNC_SPOKEN COCA_SPOKEN	
コア域	BNC_NEWSPAPER COCA_NEWSPAPER BNC_MAGAZINE COCA_MAGAZINE	基本前置詞(of, in, on, at)

今回の分析では、英米差とジャンル差のいずれがより前置詞頻度に影響しているかを検証するわけであるが、仮に、左右のいずれかにイギリス英語のデータ、もしくは、アメリカ英語のデータがかたまって布置されていれば英米差がより重要な要因になっていることを示し、逆に、英米差ではなく、同一ジャンルのものがかたまって布置されていればジャンル差がより重要な要因であることになる。この点をふまえ図 3 を確認すると、英米間でかたまっているのではなく、ジャンル間でかたまっていることが示されている。このことは、ジャンル差が英米差よりも重要であるということが統計的データから証明されたという点できわめて重要な価値を持つだろう。さらに、ジャンル差についても、一般的なジャンル分析における 5 種のテキストタイプのうち、NEWSPAPER と MAGAZINE はほぼ同じ位置に布置しており、前置詞使用に関しては、これら 2 種のジャンルにおける差異はあまりないことが示唆され、このことは、前置詞研究を進めるにあたって重要な知見であると考えられる。

上記の点をふまえて、第一象限、第二象限、第三象限、コア域の 4 つのグループの前置詞の特徴を見ていく。

まず、第一象限(右上)に布置された小説系テキストを特徴付ける前置詞については、beneath や beside などの詳細な位置関係を示すための前置詞や、thru(=through)や cept(=except)などの省略前置詞が多く見られた。以下、その用例を示す。

- (6) a. She leant against him, her head turned to one side beneath his chin, her arms tight round him. (BNC_FICTION)
- b. He picks up his briefcase, gives Frank a nasty look, then exits thru the back door. (COCA_FICTION)

(6a)のような位置を表わす前置詞については、小説では文字情報だけで読み手に場面を想定さ

せることが重要であり、これらの細かな位置関係を表す前置詞を用いて場面を描写していると考えられ、(6b)のような省略前置詞については、インフォーマルなテキストで使用されることがほとんどであることが再確認された。

つづいて、第二象限(左上)に布置された、学術的なテキストとの関係が強い前置詞の特徴として、群前置詞と動詞に由来する前置詞が挙げられる。

- (7) a. In comparison with market processes, Hayek emphasizes, political decisions are inevitably more centralized... (COCA_ACADEMIC)
- b. It seems that he would have come to the same conclusion regarding the implication of such control. (BNC_ACADEMIC)

学術的なテキストにおいては、多様な意味に解釈できるような多義語をできるだけ使用せず、意味が限定的な群前置詞や動詞由来の前置詞を用いて明確に事柄を記述しようとするため、このような表現が特徴に見られたと考えられる。

次に、第四象限(右下)に注目すると、話し言葉が布置されている一方で、vice 以外の前置詞は布置されていないことが確認される。RQ2 において、話し言葉では前置詞全体の頻度が低く、また、できるだけ単純な前置詞が使用されていたことが示唆された。この結果と合わせて考えると、話し言葉のテキストでは特徴な前置詞の使用がないことがある種の特徴になっていると考えられる。

最後に、コア域に布置された前置詞については、どのジャンルでも汎用的に使用される一次前置詞が多く、雑誌や新聞においてはあまり特徴的な前置詞使用がされないことが示唆された。

4.4 RQ4 重要前置詞の選定

RQ1~3 では、テキストタイプが前置詞使用にどのような影響を与えるかについて見てきたが、ここでは、テキストタイプの影響を受けず安定して高頻度で出現する重要前置詞を特定することを目指す。主成分分析を行ったところ、10 種の変数ごとの第 1 主成分の負荷量は以下ようになった。

表 14 主成分負荷量

テキストタイプ	第 1 主成分
BNC_SPOKEN	0.78
BNC_FICTION	0.65
BNC_MAGAZINE	0.93
BNC_NEWSPAPER	0.95
BNC_ACADEMIC	0.84
COOCA_SPOKEN	0.82
COCA_FICTION	0.57

COCA_MAGAZINE	0.93
COCA_NEWSPAPER	0.91
COCA_ACADEMIC	0.88

表 14 に明らかなように、第 1 主成分の負荷量は 10 種の変数すべて正に寄与しており、第 1 主成分を全 10 変数の総合指標とみなすことの妥当性が確認された。また、第 1 主成分の寄与率は、98.71%であり、もとのデータが持つ分散のほぼすべてが第 1 主成分に集約できていることが確認された。

上記の点をふまえ、209 種の前置詞を第 1 主成分得点の高い順に並び変えた。以下は、そのうちの上位 20 語を取り出したものである。なお、下表の「頻度率」は、該当前置詞頻度から前置詞全体頻度を除算したものである。

表 15 重要前置詞トップ 20

順位	前置詞	頻度率	主成分得点	順位	前置詞	頻度率	主成分得点
1	of	24.05%	6579130	11	about	1.19%	404309
2	in	15.10%	4359980	12	into	1.50%	271815
3	to	8.44%	2317737	13	like	0.97%	205461
4	for	8.26%	1987012	14	through	0.65%	103582
5	with	6.95%	1579349	15	after	0.75%	83879
6	on	6.30%	1439043	16	over	0.70%	72193
7	at	4.41%	971418	17	between	0.58%	67660
8	from	4.43%	906874	18	out	0.42%	53264
9	by	4.61%	839108	19	during	0.39%	49653
10	as	2.01%	425816	20	against	0.42%	6104

上表から 2 点のことが言える。

まず、1 点目は、前置詞のタイプに偏りがあることである。小西(1976)の分析枠組みでは、前置詞の大部分は「一次前置詞」によって構成されていると述べていたが、主成分分析の結果、of や in, to, for などの一次前置詞が上位 9 位までを占めていることが確認され、小西(1976)の結果を支持するものとなった。その他の前置詞タイプについては、about や into といった複合前置詞、あるいは、as や like といった他品詞由来前置詞については重要前置詞の一部に含まれているものの、群前置詞については上位 20 語に限れば抽出されておらず、もっとも主成分得点の高い群前置詞は 25 位の such_as であった。群前置詞の種別数は 88 種と、4 つの前置詞タイプの中でもっとも多かったものの、総語数の観点から見れば、重要度は低いことが明らかになった。

つづいて、2 点目は、これらの 20 種前置詞が前置詞全体の頻度の大部分を占めていることである。頻度率に注目すれば、of や in の頻度はきわめて高く、この 2 語だけで前置詞全体の 4 割近く

を占めていることがわかる。また、上位 10 語までで 84.54%、上位 20 位までで 92.11%と、これら 20 種の前置詞で前置詞全体の頻度の大部分を占めていることが確認された。この結果から、209 種の前置詞が BNC や COCA から抽出されたが、頻度について言えば、上位 20 語までの前置詞でそれらの大部分を説明できることが示唆され、英語教育においても、これらの前置詞を優先的に学習することでより効率的に前置詞を使用できるようになると考えられる。

5. まとめ

以上、本研究では、英米差とジャンル差といったテキストタイプの違いによって、前置詞の頻度が異なっているかどうか検証し、また、これらの影響をあまり受けずに安定して高頻度に出現する重要前置詞の抽出を試みた。分析の結果、以下の 4 つのことが言える。

まず、英米間の前置詞使用の違いについては、頻度上わずかな差は確認されたものの、その差は統計的に見ればあまり大きなものであるとは言えないことが明らかになった。また、イギリス英語とアメリカ英語それぞれにおいて高頻度に出現する前置詞に関しても、ほとんどが一致しており、前置詞頻度の違いはあまり大きなものではないことが示唆された。

つづいて、ジャンル差における前置詞使用の差異については、ACADEMIC > MAGAZINE > NEWSPAPER > FICTION > SPOKEN の順に前置詞の出現頻度が低くなっていることが確認され、特に、フォーマルなテキストにおいて前置詞が多用されることが明らかになった。前置詞をさらに個別的に高頻度なものに限って見たところ、上位 3 語については一致していたものの、SPOKEN において about が多用されていたり、ACADEMIC において at があまり使用されていなかったりと、いくつかの違いが確認された。

3 点目は、英米差よりもジャンル差の方がより前置詞の頻度に影響を与えていることが明らかになったことである。また、特に、ACADEMIC と、NEWSPAPER/MAGAZINE、FICTION、SPOKEN の 4 種によって前置詞の使用は大きく異なっており、ACADEMIC においては群前置詞や動詞に由来する前置詞が、FICTION においては位置関係や方向を表わす前置詞や省略前置詞が特徴的に見られた。一般的には、テキストタイプを考慮する場合、英米 2 種とジャンル 5 種の合計 10 種のグループに分けて議論を行う必要があると考えられているが、前置詞の使用に関しては ACADEMIC と、NEWSPAPER/MAGAZINE、FICTION、SPOKEN の 4 種のグループに分けることで網羅的に前置詞使用の違いを議論することが可能であることが明らかになった。

4 点目は、様々なテキストタイプにおいて安定的に高頻度で使用する重要前置詞 20 語が選定されたことである。また、それらの 20 種前置詞の出現割合は前置詞全体の出現割合の 92%以上を占めており、これら 20 種の前置詞の振る舞いを観察することで、およそその前置詞の出現パターンを解明することができると考えられ、また、前置詞を学習する上で重要な語が特定されたとと言える。

引用文献

相澤一美・石川慎一郎・村田年・磯達夫・上村俊彦・小川貴宏・清水伸一・杉森直樹・羽井左昭彦・望月正道. (2005). 『JACET8000 英単語「大学英語教育学会基本語リスト」

- に基づく』東京：桐原書店.
- 大学英語教育学会基本語改訂特別委員会. (編) (2016). 『大学英語教育学会基本語リスト新 JACET8000』東京：桐原書店.
- Francis, W., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Harlow, England: Longman.
- 小林雄一郎. (2015). 「コーパス言語学における頻度差の検定と効果量」『メソドロジー研究会報告論集』(外国語教育メディア学会)6, 85-95.
- 小林雄一郎. (2017). 『Rによるやさしいテキストマイニング』東京：オーム社.
- 小西友七. (1976). 『英語の前置詞』東京：大修館書店.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Harlow, England: Longman.
- Mindt, D., & Weber, C. (1989). Prepositions in American and British English. *World Englishes*, 8(2), 229-238.
- 竹内理・水本篤. (2012). 『外国語教育研究ハンドブックー研究手法のより良い理解のために』東京：松柏社.