



リレー連載

実践で学ぶ コーパス活用術

22

堀田 隆一

コーパスで探る英語の英米差

— 基礎編 —

© (WT-shared) Matthew 6476

この連載ではいくつかの代表的な英語コーパスを用いて様々な分析方法を紹介してきましたが、英語の変種を比較する機会はありませんでした。今回(基礎編)と次回(実践編)は、これまでの記事で取り上げてきたコーパスを含めた複数のコーパスを併用し、英語の英米差を探る方法を紹介します。くわえて、英語の英米差を、単に共時的な言語使用の違いとしてとらえるのではなく、通時的な言語変化を映し出すものとして、また一方の変種から他方の変種への言語的影響を示唆するものとしてとらえる視点を提供したいと思います。基礎編の今回は、コーパスにより英語の英米差を探るにあたっての下準備を調べましょう。

1 英語の英米差

アメリカ英語とイギリス英語は、現代世界における英語の二大変種として、日本でも広く英語学習の対象となっています。しばしば両変種の間には多くの言語的な差異があると言われますが、実際のところ、違いはそれほど大きくありません。異なる二つの変種ですから、一方に存在する表現が他方に存在しないというケースは確かに少なくありませんが、変種間の差は、同じ表現の頻度や用法の違いとして現われることが多いのです。例えば、「秋」を表す単語はアメリカ英語では fall、イギリス英語では autumn であるとよく言われますが、実際にはアメリカ英語でも詩的な響きをもつ語として autumn は使われます。また、イギリス英語でも fall はかつて使用されていました。現在でも積極的に使われることこそなくとも、用いられたところで誤解を招くことはありません。つまり、各変種における使用・不使用という問題ではなく、両変種間での頻度や用法の違いの問題と見るほうが適切な場合が多いのです。

英米差が頻度や用法の違いとして現われることが多いのは、歴史的諸事情によります。一つは、現代アメリカ英語と現代イギリス英語は、1600年頃に話されていたイギリス変種に共通の基盤をもつことです。それ以前から「秋」を表す語として autumn と fall(とその他にもいくつか)がイギリス変種に共存しており、その共存状態がアメリカにも持ち越されたのですが、その後の歴史において、イギリスでは fall は廃用に帰し、もっぱら autumn を採用したのに対して、アメリカでは原則として fall を採用しつつも autumn も詩的な響きをもつ異形として保持しました。このように、近代以降に語の採用の方向性を違えたという事情があります。

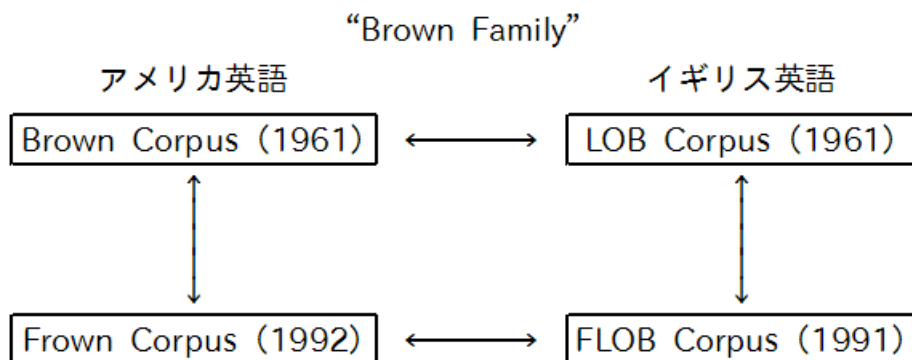
二つ目に注意したいのは、アメリカ英語とイギリス英語を比較すると言うときには、通常、アメリカ標準英語とイギリス標準英語との比較を前提としていますが、実際にはそれぞれの内部にも様々な変種があるということです。^[1] イギリスからアメリカへの移民の歴史を参照すると明らかになるように、アメリカ北東部方言とイギリス標準英語とは著しい類似性がありますし、逆にイギリス北部方言やアイルランド英語とアメリカ標準英語ともよく似ています。つまり、アメリカ英語やイギリス英語と一口に言っても、イギリス英語的なアメリカ英語もあれば、その逆もあるといったように、発音、語彙、文法などの言語項目において互いに乗り入れをしていることも少なくありません。

三つ目に、とりわけ20世紀の後半以降、イギリス英語のアメリカ英語化が進行してきているという状況があります。アクセント位置、語彙、文法などにおいて、イギリス英語はここ数十年の間にアメリカ英語の影響を受け、徐々に変化してきました。アメリカ英語化の潮流は、イギリス英語に限らず世界的に見られる現象ですが、これによって両変種の差はある程度アメリカ英語側へと縮まってきているとも言えます。

上記の歴史的事情により、アメリカ英語とイギリス英語の差異は、単純に特定表現の使用・不使用として見るのではなく、その使用頻度の違いとして見るほうが往々にして妥当ということになります。

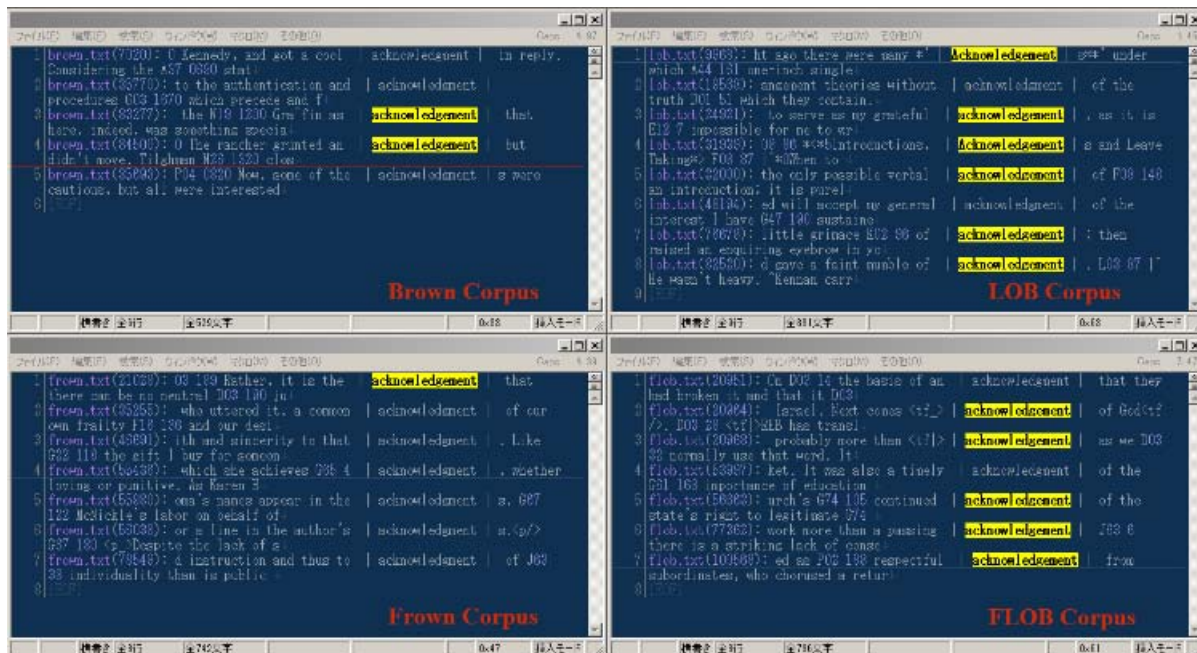
2 Brown Corpus とその仲間たち

英語の英米差が同じ表現の頻度の違いとして現われることが多いということは、比較対照するのにコーパスの利用が有効だろうということが想像されます。実際のところ、現代における英語コーパスの第一号として、1960年代初頭のアメリカ英語の書き言葉をジャンル別に計約100万語収集した Brown Corpus が現われると、それと比較可能なイギリス英語版の LOB Corpus も編纂されることになり、コーパスによる英米差の調査が可能となりました。^[2] その後、30年ほどの時間をおいて1990年代初頭の米英それぞれの書き言葉を代表する Frown Corpus と FLOB Corpus が編纂され、下図のように英語の英米差を共時的にも通時的にも調べられる “Brown Family” と呼ばれるコーパス群による研究環境が整いました。



4つのコーパスはいずれもほぼ同じ規模で、同じテキストタイプを含む均衡コーパスです。具体的にはそれぞれ約2,000語を含む500のテキストが格納されています。後にはこのコーパス・デザインにもとづいて英米以外の英語変種のコーパスも現われてきました。^[3] 近年、数億語規模の巨大コーパスが林立するなかで、“Brown Family” は各約100万語と小型ではありますが、互いに比較しやすい点が最大の特徴となっていますし、ある程度の頻度の高さを示す語彙や文法項目であれば、他のコーパスと併用しながら、今後も十分に有用であり続けられると思われます。^[4]

“Brown Family” にはウェブ上などでの使いやすいインターフェースが用意されていませんので、残念ながらこの記事で実践的な使用例を示すことはできませんが、基本的にはプレーンテキストの電子ファイルから構成されていますので、検索機能を備えた任意のソフトウェアを用いて簡単に検索できます。以下は、アメリカ英語の綴字とされる <acknowledgment> とイギリス英語の綴字とされる e を含む <acknowledgement> を4コーパスで検索した結果のコンコードンスラインを一覧したものです。後者の綴字を色づけしていますが、比較の雰囲気はわかるとおもいます。



これまでに“Brown Family”を用いて英米差に関する数々の研究がなされてきました。一つだけ紹介すると、Leech, Hundt, Mair, and Smith による英米各変種における法助動詞の頻度の推移に関する研究があります。いずれの変種においても、1960年代初頭から1990年代初頭にかけて法助動詞の頻度が下がっていることが明らかになりました。30年間での頻度減少について個別に見ると、would, will, may, should, must, shall, ought (to) は統計的に $p < 0.001$ の水準で非常に強い有意差を示し、might, need(n't) は $p < 0.01$ の水準で強い有意差を示しました。^[5] これは英米両変種をひっくりめた結果ですが、変種ごとに分けて調査すると、アメリカ英語のほうがイギリス英語よりも減少の度合いが強く、あたかもアメリカ英語の傾向をイギリス英語が遅れて追いかけているかのような分布を示しています。英米差のある語法の共時的な使用・不使用とらえているだけでは見えてこなかった細かな差異が、複数のコーパスによって明らかにされた好例と言えます。

3 Google Books Ngram Viewer

“Brown Family”は上記のようにすぐれたコーパス群ですが、簡便なインターフェースが用意されておらず、また使用目的によっては規模が小さすぎるという弱点があります。現時点で、その対極にあると言ってよいのが Google Books Ngram Viewer です。^[6] ウェブ上で極めて簡便なインターフェースが用意されており、かつ超巨大規模のデータベースとして注目に値します。2004年以來1,500万冊の様々な言語で書かれた本をデジタル化してきた Google が、そのサブセットとなる520万冊の本、5,000億語をコーパス化したものです。英語のほかフランス語、ドイツ語、ロシア語、スペイン語、中国語が含まれていますが、英語コーパスとしては British English, American English, English, English Fiction, English One Million からサブコーパスを選択できる仕様となっています。最大の特徴は、指定した5語までの検索語の頻度を過去5世紀(1500-2008年)にわたって追跡し、グラフで一覧表示してくれることです。規模が巨大すぎることも、また例文なども引き出すことができないことから、扱いやすさや厳密にコーパスとみなしてよいかどうかという点において、評価は慎重になされるべきですが、頻度を得るなどのおおまかな傾向を探るには簡便極まりないツールです。

例えば、burn の過去・過去分詞形としてアメリカ英語では burned, イギリス英語では burnt を用いられると言われますが、実際のところはどうのような分布を示すのでしょうか。Google Books Ngram Viewer でアメリカ英語の分布とイギリス英語の分布を見比べてみると、現在ではイギリス英語でも burned の使用が増えてきていることがわかります。Google Books Ngram Viewer ではこのように通時的な推移も合わせて観察することができ、英米差の問題に対して、より動的な視点から迫ることができます。

もう一つ、次回の実践編で詳しく取り上げる予定の gorgeous という語について Google Books Ngram Viewer で分布を見ておきましょう。以下の二つのグラフは、それぞれアメリカ英語とイギリス英語で gorgeous を検索した結果です(縦軸の目盛りの取り方が2:1となっており、後者が過大評価されているように見えますので、読み取りには注意してください)。

アメリカ英語における gorgeous の頻度の推移



イギリス英語における gorgeous の頻度の推移



グラフによると、この語はいずれの変種においても19世紀にはある程度の頻度で用いられていましたが、20世紀には落ち目となったようです。ところが、アメリカ英語で1980年代以降、再び勢いを盛り返してきている状況がわかります。イギリス英語の状況については判断するのは難しいですが、現在やはり復調の兆しがあるかのように見えます。この問題については、次回、別のコーパスを用いて詳しく調べていきます。

4 まとめ

今回は、基礎編として英語の英米差とは何かについて考え、英米変種間の比較を可能にするツールとして“Brown Family”のコーパス群と Google Books Ngram Viewer を紹介しました。英語の英米差は、ある表現の有無というよりは表現の頻度や用法の差ととらえるのが適切だということ、そしてそれを調べるために種々のコーパスが活用できることを示しました。“Brown Family”は英米比較にはうってつけのコーパス群ではありますが簡便なインターフェースが用意されておらず規模も小さいこと、一方 Google Books Ngram Viewer は抜群のインターフェースを備えた超巨大規模のデータベースではありますが、例文などを取り出すことができずコーパスと呼べるのか問題であることにも触れました。

今回は実践編として、“Brown Family”や Google Books Ngram Viewer の限界を乗り越えるべく、現代アメリカ英語を代表する COCA と現代イギリス英語を代表する BNCweb を用いながら、英語の英米差に具体的に迫っていきたいと思います。

〈参考文献〉

- Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith. *Change in Contemporary English: A Grammatical Study*. Cambridge: CUP, 2009.
- Leech, Geoffrey, and Nicholas Smith. “Recent Grammatical Change in Written English 1961–1992: Some Preliminary Findings of a Comparison of American with British English.” *The Changing Face of Corpus Linguistics*. Ed. Antoinette Renouf and Andrew Kehoe. Amsterdam: Rodopi, 2006. 185–204.
- 齊藤俊雄・中村純作・赤野一郎(編)『英語コーパス言語学——基礎と実践[改訂新版]』研究社、2005年。

〈著者紹介〉

堀田 隆一(ほった りゅういち)

慶應義塾大学文学部教授。2005年、英国グラスゴー大学にて Ph.D. を取得。専門は英語史、歴史言語学。英語史の記述と説明、とりわけ中英語の形態論の発達に関心がある。また、英語史の知見を活かし、いかに現代英語に関する素朴な疑問に答えられるか検討し続けており、2009年より毎日「hellog～英語史ブログ」を執筆中。著書に『英語史で解きほぐす英語の誤解——納得して英語を学ぶために』(中央大学出版部、2011年)、*The Development of the Nominal Plural Forms in Early Middle English*(ひつじ書房、2009年)がある。その他、『ライトハウス英和辞典』(研究社)などの辞書の執筆に参加している。

〈注〉

- [1] アメリカ英語にもイギリス英語にも地域方言が多々ありますし、社会方言も様々な存在します。標準英語に明確な定義はありませんが、ここではそれぞれの国で広く出版に用いられ、学校で教えられ、非母語話者に学習される英語を指すものと理解しておくことにします。
- [2] ここで触れられている“Brown Family”の各コーパスは、ICAME より入手可能です。各コーパスの概要については、CoRD の説明を参照してください。Brown Corpus よりさらに30年遡った1930年代初頭の対応する B-Brown Corpus なども編纂されています。
- [3] 現在 ICE (International Corpus of English) のプロジェクトが進行中で、英米のほか、カナダ英語やインド英語を含む14変種のコーパスが入手可能となっています。
- [4] 小型の“Brown Family”でもまだいろいろな調査が可能であるということは、Leech, Hundt, Mair, and Smith のいくつかの研究でも確認されています。ただし、比較のために注意すべきことがあるのも事実です。詳しくは Leech and Smith(186-87)を参照してください。
- [5] p 値が小さいほど有意差の検定結果の信頼度が高く、慣習的には $p < 0.01$ の水準で十分に信頼できるとされます。統計的な検定については、連載の第5, 6回目を参照してください。
- [6] Google による公式な説明はこちらの記事を参照してください。

関連書籍

『<コーパス活用> 英語基本語を使いこなす ——[形容詞・副詞編]』
『<コーパス活用> 英語基本語を使いこなす ——[動詞・助動詞編]』
『英語コーパス言語学——基礎と実践[改訂新版]』
『ライトハウス英和辞典 第6版』

キーワードで書籍検索

コーパス corpus 英語史 歴史言語学 中英語 形態論

▲ページトップに戻る

複写について | プライバシーポリシー | お問い合わせ

Copyright(C) Kenkyusha Co., Ltd. All Rights Reserved.