

コーパスを英語教育の現場に

赤野 一郎

1. はじめに

投野由紀夫氏(明海大学)が講師を務めておられるNHKテレビの英会話番組「100語でスタート!英会話」のテキストはコーパスに基づいて作成されており、この番組のおかげで「コーパス」という言葉が広く知られるようになった。しかしその実態は?その使い方は?その有効性は?ということになると、あまり知られていないのが実情ではないだろうか。本稿では、インターネット上で公開されている情報資源を基に、主要なコーパスの概要とその有効性と活用法について述べてみたい。

2. コーパスとは何か

*COBUILD Advanced Learner's Dictionary*によれば、corpusは“a large collection of written or spoken texts that is used for language research”と定義されており、言語研究を目的に収集された言語資料を指す。しかし最近では、言語分析のために一定の基準に基づき組織的かつ大規模に収集され、コンピュータにより電子化されたテキストの集積を「コーパス」と呼んでいる。以下に代表的なコーパスを紹介する。

- New ICAME Corpus Collection on CD-ROM
(<http://nora.hd.uib.no/icame.html>)

書き言葉と話し言葉のコーパス、アメリカ英語(Brown Corpus)、イギリス英語(LOB Corpus)、オーストラリア英語などの変種コーパス、時代別にサンプルを抽出した通時的コーパス(Helsinki Corpus)など、21のコーパスがある。

- British National Corpus

(<http://info.ox.ac.uk/bnc/>)

1億語のイギリス英語のコーパスで、書き言葉と話し言葉が9対1の割合で取められている。同サイトからSimple Search of BNC-World (<http://sara.natcorp.ox.ac.uk/lookup.html>)に入れば、

検索語を含む例文が出典つきで50例まで得られる。

- Bank of English

(<http://www.titania.bham.ac.uk/>)

総語数5億を越え、イギリス英語が70%、アメリカ英語が25%、残り5%はオーストラリアやカナダ英語など。そのうちの5,600万語のコーパスにアクセスできるWordbanksOnline (<http://www.collins.co.uk/books.aspx?group=155>)がある。

- 小学館コーパスネットワーク

(<http://www.corpora.jp/>)

BNCとWordbanksOnlineの日本人向けオンラインサービス。インターネットとInternet Explorerを使える環境であれば、どこからでもアクセスできる。年間使用料金は1IDにつき、BNCが15,000円、WordbanksOnlineが20,000円。

3. コーパスで何ができるか

コーパス利用の醍醐味は、コンコーダンサーという検索ソフトを使い、語と語の慣習的結合傾向、すなわちコロケーションが詳細に分析できることである。コンコーダンサーとは、検索語を行の中央に配置しその左右に文脈を添えるKWIC (Key Word In Context) 形式で検索結果を表示するプログラムである。このプログラムは検索語の左か右の特定の位置の語をキーにして、各行をアルファベット順に並べ替える「ソート」という機能を備えている。図1を参照されたい。effectの左1語目、左2語目、右1語目をそれぞれ第1優先キー、第2優先キー、第3優先キーに指定しソートしている過程を示している。ソートを行うことにより、同じ形のものが近辺に集まり、語の典型的使用の型が浮かび上がってくる。“What the computer does is to makes the invisible visible.”(Kettemann, 1995)は至言である。effectがprofoundと、さらにprofound effectがa, have および on/upon と慣習的に結合する様

1 specifically directed to obtaining this effect. Began to see last night, he wrote
 2 last night, he wrote, that duck-rabbit effect only variant on boxing-match. I had
 3 called "Hoppit" but to no effect. It is not an offence in itself that
 993 rling, will also have a corresponding effect on the sterling liquidity of the U
 994 quidity of the UK banking system. The effect of adversary politics in Britain is
 995 I see Section 16.5), the interaction effect discussed in Section 15.2.8 will be
 996 earth can be employed to eliminate the effect of the capacitances to earth on the

↓ sort 1L

1 s and localities where it takes actual effect "on the ground" (91)
 2 or causing carnage is great the actual effect can sometimes be disappointing. For
 3 ro. There are no studies on the acute effect of alcohol intake on gastric acid
 542 ne can be used to decompose this price effect into an income effect 13 (reducin
 543 ticularly gloomy stage. The principal effect would be due to their interaction
 544 ater work. They could have a profound effect upon the motor industry worldwide
 545 self, and I can remember the profound effect it had on him. These, if they pr
 546 new VAT legislation has had a profound effect on computer systems used by publis
 547 r choice will not only have a profound effect on your future life but also you w

↓ sort 2L

542 ne can be used to decompose this price effect into an income effect 13 (reducin
 543 ticularly gloomy stage. The principal effect would be due to their interaction
 544 ater work. They could have a profound effect upon the motor industry worldwide
 545 new VAT legislation has had a profound effect on computer systems used by publis
 546 r choice will not only have a profound effect on your future life but also you w
 547 self, and I can remember the profound effect it had on him. These, if they pr

↓ sort 1R

542 ne can be used to decompose this price effect into an income effect 13 (reducin
 543 ticularly gloomy stage. The principal effect would be due to their interaction
 544 new VAT legislation has had a profound effect on computer systems used by publis
 545 r choice will not only have a profound effect on your future life but also you w
 546 ater work. They could have a profound effect upon the motor industry worldwide
 547 self, and I can remember the profound effect it had on him. These, if they pr

図1 effectをソートしている過程

が、まさに見えてくる。

語がもつコロケーションに着目すれば、類義語の意味分析がより精緻なものとなる。例えば、ownとpossessの場合には、両者の意味的相違はそれぞれが目的語にとる名詞の分布の差となって現れる。(1)(2)はWordbanksOnlineでこれらの語を検索した結果得られた500例の一部である(紙面の制約と語の全体的傾向を反映させるため不要な例を削除した)。

(1)

he Netherlands population own a car, and more than 13
 ew York? "My uncle owns a company with many bar
 on concerned Mr Jones who owned a factory employing 10
 are poorer and who do not own a house, to join the ra
 New York where her father owns a restaurant. Grigory
 oss that he would one day own a store so big that the

(2)

icult to exclude people, possess a further characteri
 egments are excluded. He possessed a genius for frien
 -proportioned and usually possessed a natural sense of
 had never assumed her to possess a quality of mind li
 demanded respect. Birds possess a similar skill as y
 lations means consultants possess a talent for relatin

極めて少数の例にもかかわらず、それぞれの語の右側に来る名詞に注目すると、際立った差があることに気づく。ownは、car, company, factory, house, restaurant, storeなど、財産とみなし得る名詞を従えているのに対し、possessはcharacteristic, genius, sense, quality, skill, talentなど、大部分は当該人物を特徴づける資性を表す名詞であることがわかる。

さらにKWICコンコーダンスは、表現の微妙な意味合いをあぶり出すのにも威力を発揮する。例えば、sit throughの検索結果をご覧ください。

(3)

stice Field. I had to sit through this lecture. Apparen
 came over to Oxford, sat through this boring service,
 red. So Hetty had to sit through all those awful, awfu
 y and confused. 'I've sat through as many boring old ch
 ly law solicitor. She sat through hundreds of court cas
 e she did not want to sit through a film without chocol
 who may be forced to sit through it. Cheeky, self-mock
 is innocence and have sat through days of courtroom tes
 the business. She'll sit through four-hour meetings an
 ed followers. He then sat through a mundane 90 minutes,

外的な強制力を示す have to, not want to, be forced to, 特定の長さを表す時間表現, あるいは出席回数が多い様, 出席する会合が退屈であることを明示した表現などが見られる。sit throughはその隣接する語句と響き合い, 全体として boring, uncomfortable という色合いを帯びていることがわかる。

コーパスを使えば, 語の意味と用法に関する頻度情報も容易に得ることができる。例えば, ある英和辞典によれば, 名詞 press の語義は以下のような配列になっている。

- (4) 1 押すこと 2 アイロンをかけること 3 押し進むこと 4 切迫 5 圧搾機 6a 印刷機; 出版(業) b 校正者 7 新聞 8 報道機関 9 戸棚 10 重量挙げのプレス 11 (新聞名の)…プレス

ほぼ語義の発展順に並べられている。語義どうしの関連性を示すためには, この配列は有効である。しかしながら, 現代英語における使用頻度の観点から語義を見直すと, 結果はかなり変わってくる。参考になるのが, 筆者の編纂した『ウィズダム英和辞典』である。わが国で初めてコーパスに基づき編纂された辞書だが, この辞書では, press の語義配列は現代英語の頻度を反映し, 以下のようにになっている。

- (5) 1a 報道機関 b 記者団 c 報道 2 印刷所, 出版

社; 印刷機; 印刷 3 圧搾機 4 押すこと; アイロンをかけること 5 混雑 6 切迫 7 ベンチプレス 8 大型戸棚

コーパスを活用すれば, 語法の実態を解明したり, 従来の語法説明を検証することができる。例えば, ask A <人> B <質問> に関する従来の辞書記述において, ask B <質問> of A <人> の文型は「まれ」とされていた。改まった言い方 (formal) ではあるが, 決してまれではない。(6)をご覧ください。

(6) many, so questions can be asked of the right people. (question that hasn't been asked of Maryland coach Mark of : "Answer the question asked of you, Lieutenant," J er. Some of the questions asked of registrants concern questions may be easier to ask of an anonymous voice. S - "You have a question to ask of me?" he said. 'Yes; I ad their own questions to ask of providers and their o 受身形も含めて, Bが左方向へ移動する統語環境では, ask B of A の文型のほうが普通なのである。

4. 手軽に利用できるコーパスサイトと分析ソフト

われわれ英語教師は, 生徒の書く英語が文法的に正しいかどうかはある程度判断できるが, その言い回しが英語として自然かどうかについては, 自信をもって判定できないことがよくある。その際, 手元

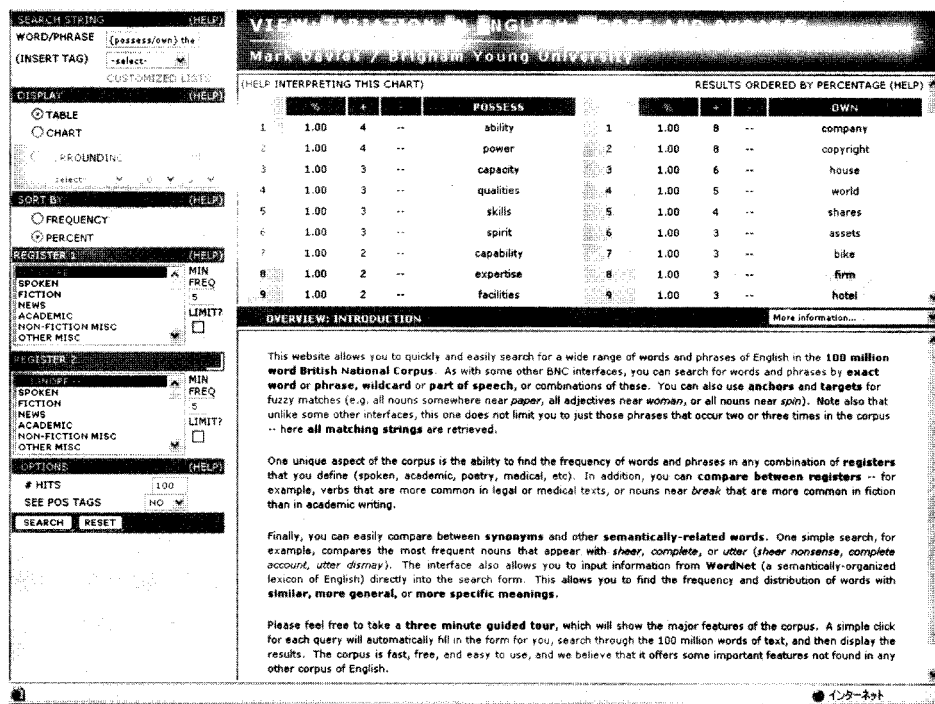


図2 own と possess のコロケーション分析

カーが身近にいるようで心強い。そこで以下にコーパスとコンコーダンサーが手軽に利用できるサイトを紹介する。

- VARIATION IN ENGLISH WORDS AND PHRASES (VIEW) (<http://view.byu.edu/>)
BNCが無料で利用できるサイト。継続して利用する場合は登録が必要。レジスター (spoken, fiction, news, academic, non-fiction, etc.) を指定してきめ細かな検索が可能。類義語のコロケーション分析が便利。例えば, (1)(2)で示したown, possessについてはほぼ同じ結果が出ている (own+ company, copyright, house, shares, assets, etc.; possess + ability, power, capacity, qualities, skills, etc.)。)
- Online Concordancer
(<http://www.lex tutor.ca/concordancers/>)
使用するコーパス (Brown, BNC (Written, Spokenそれぞれ100万語に限られている), US TV Talk, UK News, etc.), ソートの位置, 表示幅, 出力数などを指定して検索する。連語関係にある単語の頻度集計もしてくれる。
- Web Concordancer
(<http://vlc.polyu.edu.hk/concordance/>)
使用できるコーパスにAgatha Crisitie stories, Brown, LOB, The Timesなどがある。英語以外に, 中国語, フランス語, 日本語のコーパスが提供されている。
- Michigan Corpus of Academic Spoken English (MICASE)
(<http://www.hti.umich.edu/m/micase/>)
ミシガン大学での授業や面接などで交わされた話し言葉を集めた約185万語のコーパスで, 話者の性別や身分(院生・研究員など), 談話モード(講義・会話など)や談話の題材の学術分野などが詳しく記されている。

英語の教科書をコーパス化することができれば, あとは検索ソフトがあれば, 教科書の語彙分析が簡単にできる(コーパスを自前で作成する方法については, 齋藤他(2005)の3章を参照されたい)。無料のテキスト分析プログラムとしては, AntConc (<http://www.antlab.sci.waseda.ac.jp/>)がおすすめである。このプログラムはコンコーダンサーの機能のほかに,

頻度順とアルファベット順で整列した単語リストを作成する機能(Word List), 特定の単語の出現位置と出現密度を図示する機能(Plot), 指定した語数の語連結を抽出する機能(Word Cluster), 特定のテキストを特徴づけている単語を抽出する機能(Keyword List)が備わっている。Keyword Listの機能を使えば, 例えば, 英国のIndependent紙のコーパスに, 同紙の経済記事のサブコーパスをつき合わせることで, 経済分野の語彙的特徴(business, percent, market, shares, company, profits, share, rates, bid, etc.)を浮き彫りにすることができる。

Rank	Frequency	Keyword
1	837	business
2	1152	pet
3	1005	cost
4	740	market
5	531	shares
6	690	group
7	770	city
8	592	company
9	346	profits
10	323	share
11	321	uk
12	309	rates
13	254	bid

図3 経済記事を特徴づける keyword

5. おわりに

コーパスが身近になり, 利用しやすくなってきているのがおわかりいただけたと思う。自己研鑽に, あるいは教材研究に, コーパスをおおいに活用していただきたい。工夫次第でコーパスの可能性は無限に広がるであろう。

参考文献

- 井上永幸・赤野一郎編 (2003) 『ウィズダム英和辞典』三省堂
- Kettemann, B. (1995) "On the Use of Concordancing in ELT" in *TELL&CALL* (1995/4) pp. 4-15.
- 齋藤俊雄・中村純作・赤野一郎 (2005) 『改訂新版 英語コーパス言語学-基礎と実践』研究社

(京都外国語大学教授)