

はじめに

今回は、コーパスを利用した学習者英語のコロケーション分析に関して見てみることにしよう。英語で話したり書いたりする時に、外国人がいつまでもネイティブと何かが違うと感じさせる要因の1つは、適切な言葉の選び方ではないかと思う。初級学習者の場合には、基本的な英語の構造を身につけさせることが第1であろうが、中学3年から高校生にかけてはどんどん長文の英語を読むようになり、またコミュニケーション活動でもいろいろな自己表現を行なうようになっていく。やはりこの頃から徐々に、自然な単語の結びつきや頻繁に使う慣用表現などへのセンスを磨いていきたいものである。

コロケーションという言葉は、すでに英語教育でもずいぶん定着してきており、この10年くらいで英語学習辞典でもコロケーションの扱いが非常に詳しくなってきた。コーパスを利用することで、単語どうしの結びつきの度合いをうまく測定したりすることが出来るようになってきており、こういった一連の資料をもとにした教材開発はますます盛んになるであろう。

学習者コーパスを使うことで、これらのネイティブの使う自然なコロケーションに対して、学習者がよく使ってしまう不自然なコロケーションを意識させてやる効果がある。さっそく学習者データとネイティブのコロケーションを比較してみることにしよう。

コロケーション分析用ソフト

さて、実際にコーパスに基づくコロケーション分析はどのようにするのであろうか？市販のソフトでコロケーション分析がかなりよく出来るものとしては、海外ソフトでは WordSmith, TACT, 国内のものでは TXTANA であろうか。コロケーション分析と言っても、実はいろいろなレベルがある。WordSmith, TXTANA ではコロケーションの情報としては共起する単語の頻度を一覧で示してくれるだけである。図1は、TXTANA のコロケーション頻度リストの例である。これに対して、もっと強力な統計値を算出するソフトもある。TACT はコロケーション頻度のデータとして Z-score という統計量を出してくれる (Berry-Rogghe 1974)。COBUILD の Bank of English のオンライン検索ソフトやベネッセコーポレーションの辞典

WORD	TL	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5
enough	18	1	1	1	0	0	14	0	1	0	0
stress	11	0	0	0	0	0	11	0	0	0	0
hands	17	1	1	0	1	1	7	0	1	1	4
opposition	5	0	0	0	0	0	4	0	1	0	0
arm	5	1	0	0	0	0	4	0	0	0	0
point	5	0	1	0	0	0	3	0	0	0	1
one	11	1	3	1	0	1	3	0	0	2	0
men	6	0	0	0	2	0	3	1	0	0	0
feeling	4	0	0	0	0	0	3	0	1	0	0
woman	4	1	0	1	0	0	2	0	0	0	0
winds	3	0	0	0	0	1	2	0	0	0	0
support	6	0	0	1	0	0	2	2	1	0	0
stresses	2	0	0	0	0	0	2	0	0	0	0
sense	2	0	0	0	0	0	2	0	0	0	0
pressures	2	0	0	0	0	0	2	0	0	0	0
preference	3	0	0	0	0	0	2	1	0	0	0
oil	2	0	0	0	0	0	2	0	0	0	0
new	4	0	0	0	1	0	2	0	0	1	0

図1 TXTANA を用いた strong とのコロケーション頻度分布

学習者コーパスと

編集部が社内用に開発したコーパス分析ツールには t-score, MI-score といったコロケーションの強度を示す統計値が出るものもある(興味のある方は島根大学井上永幸研究室のホームページ (<http://lexis.edu.shimane-u.ac.jp/inoue/>) を参照されると、関連資料がある(その他 Church, et al. 1991 参照)。

統計値が出たほうが、その使い方さえ知っていればより効果的なコロケーションの特徴分析が出来ることは確かだ。なぜならば、コロケーションは単純にある一定のテキスト内での絶対的頻度だけでは測れない単語間の指向性があるからで、頻度が低くても特徴的に結びつく単語というものが逆に表現活動上は重要になって来たりするのである(コロケーションの分析に関する詳細は Oakes (1998)参照)。

学習者のよくやるコロケーションのミス

ここでは筆者の収集した学習者コーパス 20 万語(のべ 3000 人の英作文データ)を使ったコロケーションの抽出例を見てみよう。例えば earthquake についてのエッセイ(「大地震があったら何を持ち出すか」)が一部含まれているので、それを検索してみると特徴的なコロケーションとして表 1 のような組み合わせが出てくる。

180 例のうち、かなりが the [an] earthquake (53 例)と形容詞をつけずに用いているが、表 1 のような形容詞がついた際に、我々英語教師はどのようにコメントするだろうか? 頻度的には big が圧倒

的に多いが、big earthquake という言い方は自然なのであろうか? また興味深いのは低頻度でも我々の語感的に判断が難しそうな組み合わせ(huge, large などの形容詞の使い分けや「突然地震が来たら」という気持ちで sudden を使うなど)に学習者があえて挑戦しているところである。このような場合に、earthquake がどのような単語と結びつきが強いかをコンコーダンスなどで生徒に自分で調べさせて作文をさせれば、日本語の表現との比較をしながら、たくさんの例文に接することにもなり、非常によい動機付けになる。

ちなみに earthquake の例をネイティブのコーパスで見よう。地震という単語にヒットしやすいように、World Book Encyclopedia の本文から検索してみた(だいたいサイズは 50 万語くらい)。表 1 の右側を見ていただきたい。「大地震」という時にもっともよく来る形容詞は major であった。また severe も次いでよく使われていた。百科事典なので、文体はかなりきちんとしているから、口語ではまた違うかもしれないが、少なくとも中学生から大学生までの中で、作文で major, severe を使った生徒はひとりもいなかった。Severe あたりは我々でも気がつかないコロケーションではないだろうか。また、興味深かったことに big earthquake という言い方は World Book の本文中には出てこなかった。

Big earthquake という言い方はしないのだろうか? こういう際に、簡単にコロケーションをみるには Altavista (<http://www.altavista.com/>) の

表 1 earthquake のコロケーション頻度比較

学習者コーパス		ネイティブ・スピーカー (World Book Encyclopedia)	
全用例	180		224
big earthquake	50	major earthquake	12
great earthquake	5	severe earthquake	10
large earthquake	2	strong earthquake	6
strong earthquake	5	terrible earthquake	5
sudden earthquake	1	large earthquake	3
terrible earthquake	1	powerful earthquake	2
huge earthquake	1	destructive earthquake	2
hard earthquake	1	violent earthquake	1
disastrous earthquake	1	great earthquake	1



……(10) コロケーションの指導

words 投野由紀夫

ようなサーチエンジンで検索してみるといい。Big earthquake というフレーズで検索してみると、どうやらそういう言い方自体は可能なようだが、日本人の阪神大震災などのページでよく用いられている反面、アメリカなどでは the big one (でかいやつ) というようなフレーズで地震を指す場合が多く、big earthquake というフレーズで多用するのではないようである。How big was the earthquake? とか That was a big one. などとは言いが、big earthquake というコロケーションはそれほど高頻度ではない。日本人にはこのあたりはなかなか手強いのではないだろうか。

語法の詳細を書いてしまったが、別にこのような知識を一つ一つ持っていることが重要なのではない。むしろ、学習者が英語を使って発信する際に、コロケーションに対するセンスを徐々に身につけて行く必要があること、その際にデータを自分で見ながらその中からパタンを見つけ出して行くような data-driven learning にコーパスが非常に有効なこと、そしてさらに学習者コーパスがあれば、negative evidence として、よくやりがちな不自然なコロケーションを学習者に示してやる事が出来ること、などがポイントである。

学習辞典にもっと 発信用コロケーションを

私は学習英語辞典のユーザーサイドの研究もしているが、最近ではコロケーションを意識した辞書が増えており、これは大変いいことである。ただし、コロケーションの扱いはいくつか問題点がある。第1に、コロケーションを「日本語からの発想」で選んでいないという点である。コーパスに基づく辞書が増えて、コロケーションの記事やコラムは確かに多くなってきたが、ほとんどが英英辞典(BBIなど)を参考にしたものだ。しかし、例えば earthquake だったら、「大きい地震」は何と言うか? 「地震が起きる」「関東地方に地震が起きる」「地震は怖い」「震度5の地震」は? などなど、その単語から浮かんでくる日本語のよく使うフレーズを英和辞典の例文から探そうとすると、

あまりヒットしないことが多い。例文がどちらかということ、日本人の発想や表現にからめて作られていないからである。

97年のJACET大会のシンポジウムで、My hobby was collecting insects. という例文を紹介した(Tono 1997)。これは学習者コーパスから取ってきたものであるが、実は外国の辞書でこのような例文を載せているものは私の知る限り1冊もない。昆虫採集を趣味にするというのは、あまり欧米ではないことだからである。ところが、日本の小学生くらいではごく当たり前のことだ。そして、彼らが趣味のことを表現する際に、あってしかるべき例文だと思ふのである。

第2の問題点は、コロケーションが不足しているのはどちらかということと頻度が少し落ちた単語であり、基本2000語くらいから下の単語に情報をもっと欲しいのだが、だいたいの英語辞書は基本語重視の傾向で、コロケーションも基本語にばかり盛り込む傾向がある点である。基本語は文法的な結びつき(colligation)にかなりの紙面を割くほうがよく、語彙的なコロケーションはもっと中間層の語彙に割り当てたほうが、活用しがいのある辞書になるはずである。このへんの大きな判断をして、単語に割く紙面を従来のものよりもざっくりと減らしてしまっただろうか。

(参考文献)

- Berry-Rogge, G.L.M. (1974) Automatic identification of phrasal verbs. In Mitchell, J.L. (ed.) *Computers in the Humanities*. Edinburgh: Edinburgh University Press.
- Church, K., Gale, W., Hanks, P. & Hindle, D. (1991) Using statistics in lexical analysis. In Zernik, U. (ed.) *Lexical Acquisition: Exploring On-line Resources to Build a Lexicon*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Oakes, M. (1998) *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Tono, Y. (1997) The role of illustrative examples in English learners' dictionaries: an experimental approach. Paper presented at Symposium on Dictionary Users at JACET Annual Convention, 1997.
- (とうの・ゆきお / 元東京学芸大学講師・ランカスター大学言語学科博士課程在籍: e-mail: y.tono@lancaster.ac.uk)

学習者
コーパスと
英語指導