

1

赤須 薫

なぜコーパスか？



© (WT-shared) Matthew 6476

1 ネイティブの英語に手軽にアクセス

英語を習いはじめた頃、「今何時ですか」という疑問文は **What time is it now?** と訳す、と教わりました。ところがその後、実際の英語に接するようになると、ネイティブ・スピーカーが **What time is it now?** とは言わないことに気づきました。その代わりに耳にしたのは、**now** のつかない **What time is it?** という表現です。どちらが普通の言い方なんだろうと疑問を抱いたとき、どうしたらよいのでしょうか。ネイティブ・スピーカーに直接尋ねることができるならば、それもよいでしょう。ですが、ネイティブがいつも身近にいるとは限りません

し、母語話者が教えてくれることがすべて正しいという保証もありません。そういう時に、コーパスが役に立ちます。

それでは、コーパスに助けを求めてみましょう。**Bank of English**¹⁾ という大規模コーパスのひとつで当該表現を検索してみると、全部で**176**例見つかりました。そのうち、**now**を伴うものが**13**件でしたので、表で示すと以下のようになります：

What time is it?	163 例
What time is it now?	13 例
計	176 例

一目瞭然ですね。**What time is it?** という表現は、**now** のつかない形で用いられることが圧倒的に多いことがわかります。**What time is it now?** という言い方がないわけではありませんが、はるかに頻度が低いということを事実として指摘できます。

これは「学習者」という立場に立ってのコーパスの使い方のひとつですが、どういう表現が普通に使われているのかを知りたいと思った時にコーパスが活用できるということを示しています。このような事実を把握した上でネイティブに意見を求めると、さらによい答えが得られるやもしれません。

2 教えるべきか、教えざるべきか

「学習」と「教育」は表裏一体の問題ですが、今度は教育という観点からの活用例を考えてみたいと思います。

私はもう長いこと、映画やテレビを見ながら、これはと思う表現に出会うと「語法ノート」に記して来ました。その際、風変わりなものにばかり目が行ってしまいがちで、普通のこと／当たり前のことには無頓着になる傾向があります。一例として、**How about some coffee?** というのを聞いても、ごく当たり前の表現なので、聞き流してしまいましたが、**How about we go get some coffee?** というのに出会った時に、これはと違って一生懸命書き留めた記憶があります。**about** は前置詞ですので、このように文(正確には「節」と言います)が続くのはいわゆる「破格」な構文となります。

英語教師であれば、このような表現を教えるべきなのか／教えないほうがよいのかと思案をめぐらすところです。私自身は辞書を書く仕事をしていますので、はたしてこの表現を辞書に載せるべきか／載せざるべきかを考えねばならない立場にあります。こういう場合にもコーパスの力を借りると、問題解決の糸口が見つかります。

それでは、コーパスから採った **how about** を含む用例を覗いてみましょう(これも **Bank of English** からのものです):

sounds reasonable,” Kellman says. ‘ How about a game of chess before the M-1.Shortz: You have a good memory. How about a group with the hit ‘Red, Red
<p> Shortz: Disconcerted is correct. How about a Cabinet or Supreme Court fat than nutty chocolate brownies. And how about a Vitalinea Chocolate Orange Peter. ‘A potato bug, do you think? Or how about a quahog with its neck
(中略)

There is another scenario, sir. How about if Maguire just drove back to which you are trying to achieve. <X01> How about if we start at 15 after? <X01> in the fourth row of the orchestra. How about in the next to the last row at work: Does watering the lawn count? How about installing a hoop so you can That suggests something to me. How about it?” <p> Sidney Morton raised
(中略)

muster in the minute. Come on Chief, how about retiring on top like Michael came up with a brilliant suggestion. How about shooting the film entirely sentimental tripe. Telly for women. How about some TV drama with guts? JAY a confused pleasure/pain activity. <p> How about some of the other tricky A man’s dead and a woman’s dying. How about somebody trying to stick those
(中略)

pick your clothes up off the floor?” ‘ How about washing out the bath close enough,” the Navajo said. <p> How about writing it with my left hand?” what do we have here? <p> L. Shapiro: How about you taste my favorite, all are you doing? <p> Shortz: Not bad, how about you? <p> Sanchez: Good. Will, colleague where he would have us go. How about you? What’s your novel trying

how about の使用例は全部で **3,861** 件見つかりました。その中から無作為に **100** 例を選び、その一部だけを示しました。**how about** に続く語句には様々なものがあるのがわかりますが、(if 節を別にして)「主語＋述語」が続くパターンはここでの **100** 例中 **1** 文しか見当たりません。さらに、すべての用例 **3,861** で詳しく調べてみると、**How about we go get a drink?**

という、先に示した例文と極めて近い用例を含めて、同パターンのものが **20** 例見つかりました。このデータ内での生起率は **20/3861** ですので、**0.518 %** です。1 % にも満たないわけですから、ランダムな **100** 例中に **1** 例しか出て来なかったという事実にもうなづけると思います。

さて、これらのデータからどういうことが言えるでしょうか。まず、事実として **How about ... ?** には名詞句が続く例が大半を占めるということです。**How about you ?** のようなごく単純な名詞句もあれば、**How about writing ... ?** のように動名詞が続く場合もかなりありますが、いずれにしても、「**How about + 名詞句 ?**」のパターンが主流であり、普通であり、典型的であるということです。したがって、まず最初に教えるべき「型」はこれだと言えます。上で問題にした **How about we go get ... ?** のようなパターンは（中心から離れた）いわゆる「周辺的な」表現であることがわかります。ですので、高校生を対象とした辞書であれば、このパターンは外すという判断になるでしょう。また、一般の利用者ないし上級の学習者を対象としたものであれば、エントリーさせると思います。なぜならば、「中にはこういうパターンもありますよ」と教えることで、このような表現に出会った時に戸惑うことがないようにするほうが親切であろうと判断するからです。以上のように、「教育」という観点に立った場合にもコーパスの利用価値は同じように高いと言えます。

3 コーパス利用で何がわかるのか

コーパスを利用する価値／意義をまとめておきたいと思いま

す。そのひとつは、類似の表現をたくさん集めて観察してみることにより、そこに型／パターンが見えてくるということです。個々の表現をただ一つひとつながめているだけでは見えないものがあり、それを浮かび上がらせることが可能になってくるわけです。そして、そのパターンの中で何が一番普通に使われているのか、どれが典型的なのかを知ることができるという点も同じく、あるいはそれ以上に重要です。普通のこと、当たり前のことは意外に見逃してしまいがちです。それに気づかせてくれる、あるいは再認識させてくれるのがコーパスです。

4 コーパスとはどういうものか

話の順番が逆になりましたが、コーパスとは一般に次のように定義されます：言語を分析／研究することを目的として、実際に使われた言語をコンピューターで読み取れるような形で大量に集めたもの。早い話、たくさんの実例を収めたデータベースです。

インターネットの世界では英語が主流です。それと呼応するように、コーパスも英語のデータが中心を占めています。ただし、種類の異なるコーパスも徐々に増えてきました。最近公開された日本語の大規模コーパスもそのひとつで、今後の展望はますます明るくなったと言えます。

5 コーパスにはどういうものがあるのか

現在、利用可能なコーパスには様々な種類と相当な数があります。ここでは、今後取り上げられることになる4種類のコー

パスごとに、複雑な手続きなしに誰でもアクセスできる、無料のコーパスをごく簡単に紹介することにします。

○ 英語コーパス

BNC Simple Search: <http://www.natcorp.ox.ac.uk/using/index.xml?ID=simple>

BNC (British National Corpus) は1億語から成るイギリス英語のコーパスで、英語コーパスの中では一番よく知られているものと言っていいでしょう。**BNC** へアクセスする経路はいろいろとあるのですが、一番手っ取り早いのがこの **Simple Search** です。ただし、複雑な検索はできません。また、検索結果は1回につき50例までという制限がついています。

○ 学習者コーパス

JEFLL Corpus: <http://scn.jkn21.com/~jefll03/>

JEFLL とは **Japanese EFL Learners** (EFL は **English as a Foreign Language**) の略です。日本人の英語学習者である中学生・高校生の作文データをコーパス化したものです。

○ パラレル・コーパス (対訳コーパス)

Tanaka Corpus: <http://www.manythings.org/corpus/>

パラレル・コーパスとは、複数の言語の文と文が対訳の形でまとめられたコーパスです。利用可能な対訳コーパスはいくつかあるのですが、煩雑な手続きなしで自由にかつ簡単にアクセ

スできるものとして、これを挙げておきます。

○ 日本語コーパス

KOTONOHA Corpus: <http://www.kotonoha.gr.jp/shonagon/>

日本語のコーパスは英語のものとは比べると質と量の両面においてはるかに遅れていましたが、国立国語研究所が中心となって構築された大規模コーパスが最近公開され、利用できるようになりました。

6 おしまいに

今回のお話は以上です。イントロですので、表面をさらっとなぞっただけですが、次回以降は、それぞれの分野のエキスパートが具体的に解説をします。手取り足取りとまではいかないまでも、わかりやすさを旨として、できるだけ平易にお話しさせていただくつもりです。

この連載を読みながら、実際にコーパスをいじってみてはいかがでしょうか。水をただながめているだけでは決して泳げるようにはなりません。やはりドブンと水につかってみることが一番大切です。

〈注〉

- 1) **Bank of English** は **Collins** 社と **University of Birmingham** の共同プロジェクトにより構築され、**COBUILD** 英英辞典の編集に主に利用されているコーパス（現在、6億5千万語）である。**Collins Corpus**（現在、45億語）の一部を成すもので、一般には公開されていない。今回利用したバージョンは4億5千万語

が検索対象となっている (<http://www.titania.bham.ac.uk/>)。HarperCollins 社のコーパス・データの一部は以下のオンライン検索サービスで利用できる (いずれも有料)。

WordBanks Online (HarperCollins 社提供、5 億 5 千万語、英語インターフェース):

<http://www.collinslanguage.com/content-solutions/wordbanks>

WordbanksOnline (小学館コーパスネットワーク提供、5 千 4 百万語、日本語インターフェース):

<http://wordbanks.jkn21.com/>

〈 著 者 紹 介 〉

赤須 薫

東洋大学文学部英米文学科教授。2005-06 年、英国バーミンガム大学大学院客員研究員。専門は辞書学。日英対照、コミュニケーション研究などに興味・関心がある。『ライトハウス英和辞典 第 4, 5, 6 版』『ルミナス英和辞典 第 2 版』(いずれも研究社)の共編者。共著書に『英語辞書の比較と分析』(第 4, 5, 6 集、研究社)、*Communication in Japan and the United States* (SUNY Press)、共訳書に『学習英英辞典の歴史——パーマー、ホーンビーからコーパスの時代まで』(研究社)、『異文化に橋を架ける——効果的なコミュニケーション』(聖文社)など。