

はじめに

中 間言語の分析には、母語の影響をどうしても無視することが出来ない。とくに日本のように direct method 系がそれほど浸透しておらず、教室内で英語よりも日本語を使うことが多いような環境では、学習者が母語の影響を強く受けている可能性が高い。

学習者コーパスで、日本語の影響をどのように分析できるだろうか？ この部分はまだ研究が始まったばかりで、世界でも学習者コーパスを母語との関連で精密に研究しているグループはまだない。であるから、ここで紹介することはあくまでもまだ研究例としては基礎的なことである。しかし、今後こういった方向性の研究もあるということによって理解していただければ幸いである。

タイトルで誤解されるといけないが、この研究は学習者英語の母語からみた分析例であって、昔から行なわれてきたいわゆる日英の言語比較ではない。今回は、学習者コーパスの中に現れる日本語について分析してみる。そして次回は、筆者の温めている構想である日英対訳学習者コーパスについてそのプロトタイプを解説してみたい。

英作文の中にある日本語の分析

我々は、よく教室内でコミュニケーション活動をさせる時に、どのようにして生徒が日本語を使わないで言語活動ができるかを工夫する。しかし、実際にはなかなかそのコントロールは難しい。相当できる、真面目な生徒でないと言いたいことをすべて英語で苦勞して言うというのは無理である。

しかし、逆に考えてみると、生徒が英語にしにくい部分に焦点を当てて研究してみるというのは、非常に興味深いことだ。教師がどんなに親切に場面設定をしてやっても、どうしても生徒の伝えたい特定の語彙や表現が抜け落ちていることがある。逆に、学年やタスクが変わっても学習者がいつでも苦勞するというようなある特定の表現や構文が

あるかもしれない。我々はそういった面で、英語での表現活動の際に日本語にならざるを得ない部分にもっと注目してみるといい。

東京学芸大学で行なわれた英作文プロジェクトでは、fluency を重視する自由英作文を課すために、分からない表現があってもその部分は日本語でかまわないのでなるべくたくさん書かせるという方針を取った。その結果、生徒の作文には随所に英語に出来ずに日本語で書かれた部分がでてきた。彼らはいったいどのような部分で英語で表現することに困難を覚えたのだろうか？ 少しそのへんの詳細を見てみることにしよう。

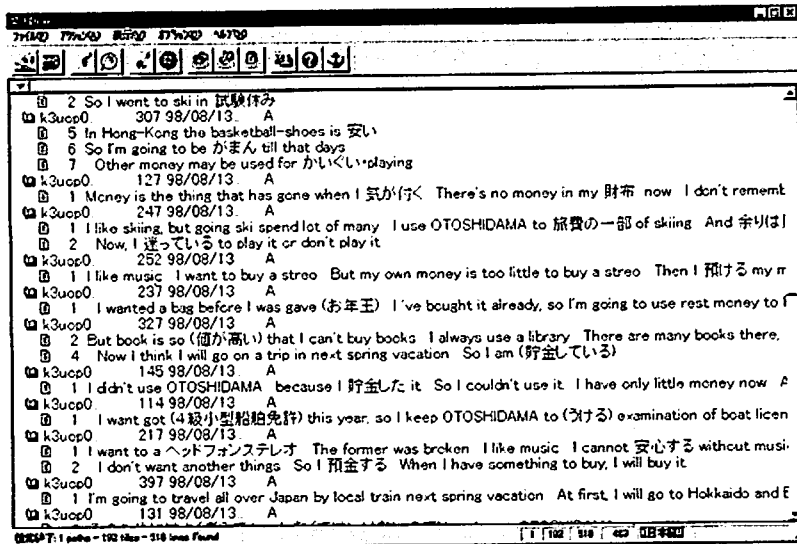
英作文中の日本語表現の抽出方法

英作文中に日本語が混在している場合、もっとも簡単なデータの処理方法はそれをそのままテキスト・エディタなどで打ちこんで GREP などのソフトで日本語部分を抽出することである。Windows 上では、Qgrep, File Diver といった高機能の GREP ソフトが利用できるし、秀丸のようなテキスト・エディタにも GREP 機能が搭載されている。次ページの図 1 は File Diver で日本語部分を取り出して閲覧している画面である。自分の書かせた英作文をこのような形で電子化して、生徒が英語で書けなかった部分を見られれば、教師にとってはこれだけでも十分に参考になるであろう。

もう少し詳細に日本語部分の分析をするためには、それなりに工夫が必要だ。1つは日本語対応のコンコーダンサーを使うという手がある。国内では浜口崇氏の作成した Corpus Wizard というコンコーダンサー (Windows 版) は日本語対応になる。また氏の作成したテキスト処理ツールの Boreal という頻度計算ソフトも日本語処理が可能だ (詳細は <http://www2d.biglobe.ne.jp/~htakashi/indexj.html> を参照)。

あるいは、日本語部分をローマ字で表記してしまうという方法がある。これによって WordSmith のような通常の英語ソフトも使用可能だ。ただし、日本語部分を抽出するために、我々のプロジェクトでは <JP>otoshidama</JP> のように日本語部分

学習者コーパスと



中学2年	中学3年
催す	集計する
預金する	目立つ
若返る	引退する
思いつく	気が動転する
仮装する	収容する
運営する	心待ちにする
見守る	
後世に伝える	
熟睡する	
昇天する	
徹夜する	
不評を買う	
力を注ぐ	

▲左: 図1 Grepソフト (File Driver) で日本語部分を抽出した画面、右: 表3 英語にしにくい動詞

中学1年生			中学2年生			中学3年生		
単語	頻度	割合	単語	頻度	割合	単語	頻度	割合
食べる	16	13%	亀	51	5%	劇	20	3%
時々	12	9%	つる	50	5%	展示	16	2%
に	10	8%	展示	24	3%	亀	17	2%
朝食	6	5%	がらくた	15	2%	演劇	11	2%
朝ご飯は	4	3%	貯金する	15	2%	早弁	8	1%
が	4	3%	演劇	12	1%	お年玉	7	1%
サラダ	4	3%	貧	8	1%	装飾	7	1%
だから	4	3%	味噌汁	8	1%	味噌汁	7	1%
ほとんど	4	3%	ステージ	7	1%	イベント	5	1%
朝	2	2%	ためる	7	1%	催し物	5	1%

◀表2 各学年の英語に出来なかった単語ベスト10

をタグで囲ってある。こうすることで日本語部分を1発で検索可能だ。

また日本語部分の検索だけでは単純に文字列しか得られないので、それを再加工してより使えるリストにするために日本語の形態素解析プログラムを使うといい。奈良先端科学技術大学院大学の松本研究室では、汎用の日本語形態素解析プログラム「茶釜(ちゃせん)」を公開している。このWindows版を用いれば、より精度の高い分析が可能になる(詳細は <http://cl.aistnara.ac.jp/lab/nlt/chasen.html> を参照)。

学年ごとの日本語使用率の比較

さて、ツール類の紹介はこのくらいにして、分析の結果を見てみよう。表1は、中学1年から3年までの自由英作文における日本語使用率を比べ

たものである。被験者は某国立大学附属中学で、データ数は1年生が作文数36件(総語数1245語)、2年生515件(2万3341語)、3年生468件(2万4209語)である。1年生のデータはまだ我々のコーパスでは少なく、十分比較の対象にならないかもしれないがご了承いただきたい。

	中学1年	中学2年	中学3年
異なり語に占める割合	29%	23%	18%
総語数に占める割合	6%	4%	3%

表1 日本語使用の割合の変化

異なり語(英語では type という)に占める割合というのは、英作文の中で出てきた異なり語の中でどのくらい日本語があったかを示すものである。中学1年生では約3割が日本語であるが、中学3年までに徐々に減少して行き、2割を切るようになる。また総語数(token)の中での使用率を見ると、中学1年では6%で、20語に1語あまり日本語を



……(7) 日英比較分析 [1]

words 投野由紀夫

使う計算になる。中学3年くらいまでには3%にまで半減し、相当英語だけでの自由英作文も可能になる。

学習者が困難を感じる日本語表現

それでは実際にどのような言葉がもっとも困難を感じるのだろうか？ この分析は少々面倒であるが、単純な日本語の文字列リストを形態素解析して見出し語にまとめ、その頻度を取りなおすという形にした。また、修飾関係や活用形などに関する情報が落ちてしまうので、見出し語にまとめる際にその variation を記述するようにした。

表2が、学年ごとのもっとも英語に出来なかった日本語のベスト10である(割合はこれらの単語の日本語で書いた単語全体に占める割合)。このリストでは明らかにトピックによる特徴的な語彙(例えば食事についてや浦島太郎の話の続きを書かせたので「亀」や「鶴」などが現れており一般的な意味での頻度リストにはなり得ないが、それでもリスト全体を注意して見て行くと面白いことがわかる。第1に、中学1年生の場合には「に」や「が」のような助詞がけっこう頻繁に出てくる。日本語の助詞の英語での表わし方がまだはっきり分からないせいか、英作文中でそれを補おうとする傾向がある(他にもリストに「から」「でも」「とは」「にも」「より」「を」などが現れており、これらは中学2,3年のリストにはないものである)。

次に中学1年の英作文では「だから」のような接続詞や「ほとんど」のような副詞表現がまだわからないので日本語になって現れることが多いが、中学2,3年では徐々にそういった表現よりも名詞や動詞の content word の不明なものを日本語化する傾向が強くなる。

中学1年の「食べる」がトップなのは意外な感じがするが、実はこの表現は形態素解析でまとめられているので、実際の作文中に出てきた表現は以下ようになる:

食べ終わる(1); 食べ始める(1); 食べます(1); 食べています(1); 食べるのは(2); 食べるものは(1); 食べられるかな(1)

このように「食べる」だけではそれほど難しさは感じないが、「食べ終わる」「食べ始める」のように動作の開始や終了を表現しようとしたり、「食べるのは」「食べるものは」のように動詞を用いたより高度な言語表現の際に、困難を感じているようだ。高学年になると、このような動詞の時制や相、準動詞の表現に関する日本語化は極端に少なくなる。なんとか自分で工夫してそれらを言おうとするようになるのであろうか。このへんは今後ももっとデータが必要な部分である。

中学2,3年になると前述したように、リストにかなり名詞が多くなる。またそれだけではなく、簡単に言えるような表現でも、日本語が難しいために対応する英語が浮かばないという現象が散見されるようになる。例えば、表3は簡単な動詞表現に言い換えれば比較的楽に英語になるようなものを難しい日本語で言ってしまう例だ。日本語の表現力がどんどん増して行く中学生の時期には、このような難解な表現を用いて英語と結び付けてしまうような習慣がつくと、英語が自然に書けなくなることが予想される。このような学習者にとって難解な英語表現に関するデータは、英作文指導、コミュニケーション活動の一環で語彙を補強させてやる場合、または教材・辞書作成などの重要な指針となるだろう。しかし、我々はこの種類の学習困難語に関して組織的なデータを持っておらず、またそれをどのように克服するかの適切な指導法も明確ではない。ここで紹介したような研究が今後ますます期待される。

今回は、学習者コーパスを用いた英語になりにくい表現についての考察を試みた。まだデータは乏しく、あまり明確な結論は出ないのだが、コーパスを活用した語彙分析に日英比較の観点を加えることで、非常に興味深い言語学習上の特徴をとらえられる可能性があることは示せたと思う。今回は、学習者コーパスの日英比較分析の方法をますます精密化するために、筆者が構築予定の日英学習者パラレルコーパスの構想とそのプロトタイプの紹介をしてみたい。

(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: email: y.tono@lancaster.ac.uk)

学習者
コーパスと
英語指導