OS-1

【学術賞受賞者講演】

英語学習語彙研究とDDLに関する研究

[英語コーパス学会学会賞（英語コーパス学会）「英語学習語彙研究とパラレルコーパスを利用したDDLに関する研究」平成20年10月4日]

日大生産工　○中條 清美

1 はじめに

英語コーパス学会は，コーパス（言語データベース）とコンピュータを用いた英語英文学の研究とその応用を目指す日本で唯一の学会です。英語コーパス学会学会賞は，コーパスを用いた英語および関連領域の研究に関する研究業績に対して授与されるものであり，本賞は「英語学習語彙研究とパラレルコーパスを利用したDDLに関する研究」を対象として授与されました。

コーパスとは「言語研究に利用できるコンピュータ処理可能な言語データの集合体」のことです。幅広いジャンルの言語テキストから言語データを体系的にサンプリングして大量に集められて構築されます。1990年代になって，１億語のBritish National Corpus (BNC)が完成し，コーパス言語学は飛躍的に発展しました。現在，コーパスの大規模化，多言語化，種類の多様化が進行し，その研究の可能性は，語法研究，文体論，語彙調査，通時的研究，外国語教育への活用等に広がっています。

コーパスが科学技術分野に活用されている実例には，ライフサイエンス分野の LSD (Life Science Dictionary) プロジェクト[1)]や科学技術分野の PERC（Professional English Research Consortium) Corpus[2)] があります。後者は科学技術・理工系22分野を代表する1,700万語の論文からなる世界最大の科学技術英語コーパスです。無料で公開されており，今後，理工系の研究者によって活用が期待されるコーパスの１つです。

外国語教育におけるコーパスの活用例には図１のように，英語の文例に対応する日本語訳が表示される日英パラレルコーパスがあります。日本語訳によって英文の難易度を下げることができます。加えて，文脈の中で，英語の実際の使用例を日本語の訳とともに，観察できるという利点があります。

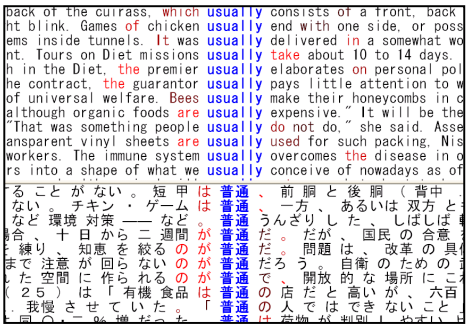

図１ “usually”の検索結果

このようなコーパス言語学の発展を背景に，講演者は，BNCやPERCなど多様なコーパスを利用した語彙調査や，多言語コーパスを活用した外国語教育の実践研究を行ってきました。以下，本稿では，受賞対象となった研究について例をあげて説明します。

2 英語学習語彙研究[3), 4), 5)]

語彙の効率的な学習・指導には「基本語彙」と「専門語彙」を峻別して指導すると効果的であると言われます。具体的には，最初に2,000語程度の汎用性の高い基本語彙を学習した後に，学習者の学問領域や職域に合わせて目標分野を限定し，その特定分野の専門語彙を指導することで学習効率を高めるというものです。

専門分野の教育用語彙選定には，当該分野に関する高度な知識が要求され，また学習者の習熟度レベルに合致した語彙の配列が必要とされるため，専門語彙の選定作業は容易ではありません。このような現状に鑑みて，簡便かつ客観的な方法で，習熟度レベル別に，精度良く，専門分野に特徴的な語彙を抽出する手段として，複数の統計指標を用いる方法を提案し，それらの指標が専門語彙を抽出する結果の有効性を検証しました。

例えば，ビジネス分野に特徴的な語を抽出した場合，表１のように各統計指標によって異なる語彙レベルの特徴語が抽出されること

がわかります。Freq，Dice，Cosine，CSMの指標が抽出する特徴語は母語話者の3～5年生レベル，LLR，Chi2，Yatesは6年生レベル，PMI，McNemarは9～10年生レベルに相当します。さらに図２より，LLR，Chi2，Yates，PMIの4指標の特徴語の約8割が専門辞書の見出し語と一致し，専門語彙が高い精度で抽出されることが検証されました。上述のような結果は複数の分野にわたる特徴語抽出に安定して現れることが確認できました。したがって，指標を適切に選択して利用することで，学習者の習熟度に応じて，初級･中級・上級といった段階的な専門語彙を選定できる可能性が示されました。

表１　9種の統計指標によって抽出されたビジネス分野特徴語の平均学年

統計指標	平均学年レベル（US grade）
Freq / Dice	3.3
Cosine	4.3
CSM	5.0
LLR	6.1
Chi2 / Yates	6.5
PMI	9.0
McNemar	10.2

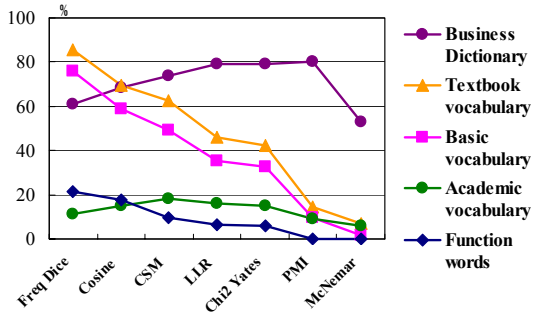

図２　9種の統計指標によって抽出されたビジネス分野特徴語の基本・専門語彙含有率

３　パラレルコーパスを利用したDDL研究[6), 7)]

コーパスは英語が実際にどのように使われているかを観察して英語のさまざまな規則性を見出す資料として実証的英語研究において活用されています。こうした活用方法は外国語教育との親和性が高く，データ駆動型学習（Data-Driven Learning: DDL）と呼ばれる学習が提案されています。DDLでは文法規則や語彙の意味・用法等を検索結果（図１）から観察して，学習者自身が発見し帰納的に学習することができます。DDLは英語学習に有効であろうと言われながら，(i) コーパステキストの難易度，(ii) インターフェース，(iii) 指導法の欠如という3つの壁があるため，教育利用に関する研究は世界的に立ち遅れています。

講演者はこれらの壁を乗り越えるために次のような方策を講じてDDL実践の研究を進めています。 (i) 日本では，ほとんど活用例がなかった日英パラレルコーパスをいち早く英語教育に取り入れ，日本語訳を利用して英文理解の難易度を下げるという方法で，英語教育でのコーパス利用を可能にしました。(ii) 教室でコーパスを利用するには検索ツールが必要ですが，現在は有料です。そこで，誰もが自由に使える無料の検索ツールを開発中です。(iii) 仮説形成から仮説検証を経てプロダクションに至る「４ステップDDL指導法」という指導手法を提案し，実践効果を報告しています。

現在，平成21年～24年度科研費を受けて，英語教育のDDL教材の開発・実践・効果検証の研究を行い，コーパスの英語教育における実践的利用の展開を目指しています。

参考文献

1) LSD (Life Science Dictionary)プロジェクト http://lsd.pharm.kyoto-u.ac.jp/ja/about/about_lsd/index.html

2) PERC（Professional English Research Consortium) Corpus http://www.corpora.jp/~perc04/index_j.html

3) Chujo, K. “Measuring Vocabulary Levels of English Textbooks and Tests Using a BNC Lemmatised High Frequency Word List”, in J. Nakamura, N. Inoue and T. Tabata (Eds.), *English Corpora under Japanese Eyes*, Rodopi, Amsterdam, (2004), pp. 231-249.

4) Chujo, K. and M. Utiyama, “Selecting Level-Specific Specialized Vocabulary Using Statistical Measures”, *System*, 34 (2), (2006), pp.255-269.

5) 中條清美，内山将夫，中村隆宏，西垣知佳子，教育用語彙選定における特徴語抽出及びWordplotの利用，言語処理学会第13回年次大会発表論文集, (2007), pp.474-477,（優秀発表賞）

6) Chujo, K., M. Utiyama and S. Miura, “Using a Japanese-English Parallel Corpus for Teaching English Vocabulary to Beginning-Level Students”, *English Corpus Studies*, 13, (2006), pp.153-172.

7) Chujo, K., M. Utiyama and C. Nishigaki, “Towards Building a Usable Corpus Collection for the ELT Classroom”, in E. Hidalgo, L. Quereda and J. Santana (Eds.), *Corpora in the Foreign Language Classroom*, Rodopi, Amsterdam, (2007), pp. 47-69.