

## 英語辞書編纂とコーパス

八木克正

(関西学院大学)

## 1. 言語研究とデータ

言語学では、コーパス (corpus.複数形はcorpora) は本来、言語資料をさす。これが近年のコンピュータの発達に伴って、コンピュータが処理可能な電子コーパスをさす用語になった。従って、コーパスというと今や電子コーパスと同義語となった。そして、その電子コーパスは英語辞典の編纂には欠かすことができない時代になった。今や、コーパスは、辞書編纂ばかりではなく、英語のあらゆる面からの研究すべてにわたって欠かすことができないツールといってよい。

もともと言語の研究は実証的なものである。すなわち、言語研究は言語事実に基づいておこなわれるべきものであることは科学的文法研究の大前提であった。この考え方は、文法書 *A New English Grammar*, 2 vols. (1881-1898) を作った H. Sweet によって明らかにされた。英語辞書についてはもっと歴史は古く、S. Johnson の *A Dictionary of the English Language*, 2 vols. (1755) は、数多くの文学作品から用例を引き帰納的に意味を分析をする科学的手法をとった (このような事情は、八木 (編著) 『新英語学概論』 (英宝社, 2007 : 77ff, 168f.) を参照されたい)。20世紀初頭の科学的伝統文法研究者である O. Jespersen, H. Poutsma, E. Kruisinga などがその伝統を発展させた。

また、アメリカ構造言語学でも、データ重視の研究が発展し英語研究についても実証的な研究がなされた。C. C. Freis の *American English Grammar* (1940) は合衆国政府所有の日常的な手紙を使っている。また同じく Fries の *The Structure of English* (1952) では「合衆国北中部の大学社会で録音された英語の話言葉」 (actual English speech observed and recorded in a university community in the North-Central part of the United States) がもとになっている。同じく構造言語学の流れをくむ Martin Joos の *The English Verb : Form and Meanings* (1964) は、Sybille Bedford (1958) の裁判記録をイギリス版とアメリカ版の両方を使って英語の動詞の意味と形式の研究を行っている。このような実証的な言語研究の手法が、生成文法の出現で根本的な転換期を迎えることになった。

## 2. データと直観

N. Chomsky が創始した生成文法の根本的な目的は、人間がもつ言語能力を解明することである。人は、創造的にいくらかでも文法的な文を生成する能力をもっている。その能力を解明するのが生成文法の役割である。人は、いわば文法的な文を生成する機械である。その機械の仕組みを解明することが生成文法が考える言語学の役割である。従って、言語資料は、その言語を話す母語話者が必要に応じていくらかでも作りだすことができる。言語研究には、他人が話したり書いたりするものを資料とする必要などないということになる。

このような考え方は、伝統的な科学文法や構造言語学の枠組みで行われた研究の方法と

根本的に対立する。生成文法の影響力が大きくなればなるほど、実証的な研究がおろそかになったのはこのような考え方の広まりと関係がある。

しかし、このような生成文法の考え方は、常に情報提供を得ることができるインフォーマントがいらない限り、日本語を母語とする我々が英語を研究することができないことはその論理的帰結である。しかし、言語研究は、人間の言語能力を解明するためだけではない。日本語を母語とする我々にとって英語と日本語を対応させるという重大な仕事がある。そのほか、言語の多様性（英語や日本語の話者の話すそれぞれの母語は多種多様である）、個人的な好みの違いなどを考えると、それぞれの母語話者が自分の言語を研究すると、研究者の数だけ異なった文法ができる可能性がある。J. Sampson (2001) は、科学の実証性の立場から、言語学者が自分の頭の中の出来事を研究対象にすると、第三者がそれを検証することができないという根本的な問題点を指摘している。

コーパスを使った言語研究によってはじめて、第三者が検証できる科学的な研究ができること、また母語以外の言語でも母語話者と基本的には対等の立場で研究ができることへの道を拓いたと考えることができる。

### 3. 英語辞書とコーパス

それでは、英英辞典と英和辞典がそれぞれに使ったと公称しているコーパスを以下にまとめておこう。左から辞書名、コーパス名、そのコーパスの語数である。中には語数が明記されていないものもある。

<i>COBUILD</i> <sup>5</sup>	the Bank of English.	645mil. words
<i>LDOCE</i> <sup>4</sup>	the Longman Corpus Network.	300mil. words.
<i>CALD</i> <sup>2</sup>	the Cambridge International Corpus.	600mil. words.
<i>OALD</i> <sup>7</sup>	British National Corpus, the Oxford Corpus Collection, the Oxford Reading Programme for language research.	
<i>MED</i>	the World English Corpus.	200mil. words.
<i>NOAD</i>	the Oxford databank (100mil words) + North American Reading Program.	87mil. words.

ジーニアス英和辞典第	独自コーパス
ウイズダム英和辞典	三省堂コーパス
ユース・プログレッシブ英和辞典	BNC, 追加のアメリカ英語コーパス
ロングマン英和辞典	Longman Corpus Network, Longman Corpus of Contemporary Japanese, 英語学習者の誤用コーパス.

### 4. 今後の展望—特に英和辞典改善のためにコーパス活用を

英語辞典を母語話者の直観だけで作ることができずが自明である。コーパスを駆使できるようになって、辞書記述にも革命的な内容の充実がはかれるようになった。特にcollocationの記述が豊かになり、語法的な情報が精密になった。だが、特に英和辞典には多くの課題がある。英和辞典の保守性の改革と新しい英語の傾向への対応である。英和辞典には驚くほど古い記述が残されていることは、筆者が長年にわたって指摘してきたところである。また、常に変化する英語の傾向を知るうえでも、今の英語を集積したコーパスを使って今phraseologyの観点からの新しい情報を英語辞書に盛り込むことは極めて重大な今日的課題である。