

爆発 する 英語



新語の「爆発」

データに見る
新語収集最新事情

投野由紀夫

Tono Yukio

新語の狩人たち

英語辞典のジャンルの中で新語が重要な位置を占めるのは、通例最も収録語数が多く網羅的な大辞典クラスである。表1は、Random House社が自社のカレッジ版辞典や大辞典に採用した、時代を代表する新語の過去60年間の変遷を表にした

1940s	1950s	1960s	1970s	1980s	1990s
A-bomb	aerospace	cable television	chairperson	compact disc	bad hair day
bikini	computerize	genetic code	diskette	emoticon	granny dumping
cheeseburger	desegregate	hippie	passive smoking	in-line skate	intranet
jet plane	hash browns	jet lag	personal computer	managed care	nanotube
quiz show	junk mail	megabyte	Silicon Valley	palmtop	superchurch
tape recorder	neutron bomb	sexism	video game	Sandwich generation	V-chip
TV	weirdo	space shuttle	X-rated	virtual reality	web site

表1：Random House社で過去60年に採用された新語

ものである。

これを見ると新語がまさしく時代を映す鏡だといことが分かる。しかし、実際は大辞典ほど改訂のサイクルが長いので、米国では新語の採録などは小回りの効くカレッジ版クラスで宣伝することの方が多い。日本の辞典市場では、『リーダーズ英和』(研究社)の一人勝ちだった携帯版大辞典と言われるジャンルに『グランドコンサイス英和』(三省堂)が参入。またポケット版も『ポケットプログレッシブ』(小学館)が78,000項目、それを追うように『エクシード』(三省堂)が12万項目、と収録語数は日増しに増えている。大英和クラスにも25万語を擁して『ジーニアス』が参入するなど、まさしく21世紀初頭の英語辞書戦争は「爆発する語彙をどう捉えるか」が1つの勝敗の分かれ目になってきている。それと同時に新語収集の方法も変化を遂げてきている。本論では、この新語の爆発の特徴とその収集方法にスポットを当て、辞書編纂の最新事情を見てみたい。

伝統的新語収集の方法

新語収集の伝統的な方法はフィールドワークである。Merriam-Websterの辞典編集者達は、1日のうちの1、2時間を読書に充てるそうで、彼らはこの時間を“reading and marking”と言っている。新聞、雑誌、広告、インターネットのwebページなどありとあらゆる書き物を読みながら、新しい単語、従来からある単語の新しい意味や語法、新しい綴り、屈折形などを注意深く探すのだそうだ。面白い単語にはマークをしておいて、3×5インチのカードとコンピュータの両方

に引用例文として記録しておく。彼らはこの慣習を1880年代からずっと継続しており、引用カードの総数は1,500万件、引用例文データベースでは総語数5,000万語が利用できるということだ。

また Random House の Wendalyn Nichols 女史いわく、「どんなに電子環境が整っても、新語だけは辞典編集者のフィールドワークが絶対に必要」とのこと。Random House は現在、急ピッチで編集体制を電子化しているが、新語収集に関してはフィールドワークを今でも重視している。

市民参加の文化事業的手法

新語収集に欠かせないもう1つの伝統的アプローチが、一般市民を巻き込むという手である。この代表格が OED である。1997年に第2版への補遺による改訂を終えた OED は、2000年3月から web によるオンラインサービスを始めた。それと同時に2010年完成に向けて第3版の改訂が開始され、去る1999年8月に編集主幹の John Simpson が大々的なキャンペーンを行い、新語・新語義の登録を一般読者から募集した。これは James Murray が1879年に英国市民に訴えて400名以上が8万件もの情報を寄せたことを髣髴とさせる新たな市民参加の呼びかけだった。その後、このプロジェクトがどの程度進捗しているのか、<http://www.oed.com> で紹介している。

インターネットが変える新語事情

おそらく上記2つの伝統的新語収集法（フィールドワークと市民への呼びかけ）は、一部の例外を除いては欧米の辞書出版社が主として採用する方法だ。日本の場合にはより簡便な方法として欧米の新辞典からの新語採集が最もオーソドックスな方法だった。新語は実際、どの程度その用法が言語的に安定した使用になっているかで、採録の判断が難しい。案外、たまたま拾った単語が個人語(idiolect)であったりすることもある。日本の英語辞書では、まず欧米の典拠辞書の確認をとって載せるという方が無難なのである。そのために、Barnhart の新語辞典などが重宝されている。

しかし、最近はこの確認方法も様変わりしてきているようだ。インターネットの10年間の爆発的な進歩によって、現在閲覧できるオンライン辞書

表2：オンライン辞書リンクの代表的サイト

- ◎ 辞書の辞書
(<http://www.monjunet.ne.jp/PT/bin/dict.dll>)
- ◎ 翻訳のためのインターネットリソース
(<http://www.kotoba.ne.jp/>)
- ◎ 辞典&翻訳の横断検索システム
(<http://www.jah.ne.jp/~takanori/dict.html>)

の数は驚異的に増えている。また、各分野の専門家が作る web ページの専門語欄なども充実しており(表2)、辞書出版社の最近の傾向としてはインターネットからの新語・専門語収集のプロ集団を作るという傾向が出てきている。これらがまだ欧米の辞書にさえ載っていない用語のチェックを web 上で行い、Altavista などの検索エンジンでの使用頻度などを確認し、定着度を判断したりできる時代になってきている。日本の英語辞書が欧米に先んじて新語を収録するということが可能な時代になってきているのだ。

インターネット上に現れた新語データバンク

それだけではない。インターネット上には、新語を専門的に収集するサイトも続々現れてきている。例えば、The Word Spy(<http://www.logophilia.com/WordSpy/>)では読者からの投稿を中心に新語のデータバンクを公開しているが、各新語データには必ず語義と1つの実際に使われた例文が出典付きで記載されており、参考になる(表3)。また、The International Dictionary of Neologism(<http://www.net22.com/neologisms/neoframe.html>)というサイトもかなりの数の新語を読者から集めているが、こちらの方は出典など今ひとつ不明確である。コンピュータ関連の専門用語では、What is.com(<http://whatis.techtarget.com/>)が非常に網羅的な情報を提供してくれる。その他、World Wide Words(<http://www.quinion.com/words/index.htm>)、Rice 大学の Suzanne Kemmer が自分の言語学の授業で作っている新語ページ

表3：The Word Spy の項目A見出し語(抜粋)

- active aging
- ad creep
- address munging
- adhocracy
- adminisphere
- adulescent
- advermation
- advertecture
- advertorial
- affective computing
- affluenza
- aggressocracy

(<http://www.owl.net.rice.edu/~ling215/NewWords/index.html>) も一見の価値がある。

自然言語処理と新語研究

新語収集に関しては、もっと大規模なプロジェクトとして取り組んでいる研究機関もある。代表的なものでは、マンハイム大学の Norbert Volz を中心とした CORDON, カナダの Queen's University of Kingston で行われている NeoloSearch, そして特に Liverpool 大学の Antoinette Renouf が中心に行っている AVIATOR, SPRING というプロジェクトが注目している。1990~93年の, AVIATOR プロジェクトでは *The Independent* 紙のコーパスから新語を自動収集するシステムを開発し, その収集結果を <http://www.rdues.liv.ac.uk/newwords.html> で公開している。

基本的には, コーパスから単語リストを作成し, それに新しい新聞記事の単語を照合していくのだが, 彼女らの研究の面白い点は, 新語造成のメカニズムを知るための1つの手段として, 新語としてリストに新たに追加されるべき単語の形態素解析を行い, どのような造語の方法が最も活発に新語を作り出すのかを調査している点だ (詳細は Renouf 1996 参照)。

表4は AVIATOR の後継である APRIL プロジェクトで1998年4月の *The Independent* 紙から新語として抽出された単語の品詞別の内訳を示したものである。固有名詞がトップで全体の約4割を占める。普通名詞(第2・4位)の35%と合わせると, ほぼ4分の3は名詞になることがわかる。また新語を作る接辞に関しても, 表5のように統計情報が得られるようになった。

このような言語統計情報が蓄積されてくると,

固有名詞	2587 (38.2%)
単数普通名詞	1711 (25.3%)
形容詞	846 (12.5%)
複数普通名詞	702 (10.4%)
動詞-ing 形	307 (4.5%)
動詞過去分詞形	305 (4.5%)
動詞原形	103 (1.5%)
副詞	91 (1.3%)
動詞過去形	68 (1.0%)
単数場所名詞	53 (0.8%)

新語収集もただ既知の単語リストと新データを比較するだけでなく, コンピュータが内部構造を解析して新語として目星をつける単語の絞り込みを行い, それらの単語に関して使用

表5：新語を作る接辞トップ10

	接頭辞	接尾辞		接頭辞	接尾辞
1	post	's	6	re	y
2	un	s	7	ex	style
3	dee	r	8	super	like
4	non	ed	9	pre	ish
5	anti	ing	10	pro	ly

されているコンテキストの解析を重点的に行うことで, より確実に新語の同定を行うことが可能になってくる。

新語収集の自動化に取り組む辞書出版社

最後に, 辞典出版社の取り組みも紹介しておこう。例えば, 辞典出版では後発の Cambridge University Press では, 新語の抽出は主として自社のコーパスを用いて行う。辞典編集長の Patrick Gillard 氏によると, 現在, CUP では社内用コーパスとして web 巡回ソフトを用いたテキスト収集を行っており, 常に数百万語単位のアップデートを定期的に行っている。Corpus manager の Andrew Harley 氏をはじめとするデータ処理チームが, 既存の見出し語リストとの照合作業を行い, レキシコグラファーに参照リストを渡す。彼らは, それをオンラインのコーパスで文脈を確認などして, 見出し語として採用するかの判断をする。CIDE の出版から *Cambridge Learner's Dictionary* という小型学習辞典を出した CUP は, 新語収集のシステムを組み上げながら, 大規模データを駆使した大型辞典の開発に取り組むのだろうか? 興味深いところである。

おわりに

以上, さまざまな辞書編纂における新語収集に関する方法論やアプローチを概観してきた。実際に新語を辞書に入れる判断は, 外国人である我々にはきわめて難しいものである。しかし, ネイティブの書いた新語辞典というフィルタだけに頼っていた時代は終わった。この世界の加速度的な変化に対応するためにも, 辞書出版社は新たな編纂体制やインフラの整備を迫られていると言えよう。参考文献: Renouf, A. and Baayen, H. (1996)

Chronicling the Times: Productive Lexical Innovations in an English Newspaper. *Language*, 72 (1): 69-96. (明海大学助教授)