

概説

コーパス言語学

手法・理論・実践

著 トニー・マケナリー

Tony McEnery

アンドリュー・ハーディー

Andrew Hardie

訳 石川 慎一郎

Shin'ichiro Ishikawa



ひつじ書房

CORPUS LINGUISTICS

概説

コーパス言語学

手法・理論・実践

著 トニー・マケナリー

Tony McEnery

アンドリュー・ハーディー

Andrew Hardie

訳 石川 慎一郎

Shir'ichiro Ishikawa



ひつじ書房

概説コーパス言語学—手法・理論・実践

Corpus Linguistics: Method, Theory and Practice

Tony McEnery & Andrew Hardie

Japanese translation by Shin'ichiro Ishikawa

発行 2014 年 5 月 30 日 初版 1 刷

定価 3800 円 + 税

著者 トニー・マケナリー、アンドリュー・ハーディー

訳者 石川慎一郎

発行者 松本功

装幀 中野豪雄 + 川瀬亜美 (株式会社中野デザイン事務所)

印刷製本所 株式会社 シナノ

発行所 株式会社 ひつじ書房

〒112-0011 東京都文京区千石 2-1-2 大和ビル 2F

Tel.03-5319-4916 Fax.03-5319-4917

郵便振替 00120-8-142852

toiawase@hituzi.co.jp <http://www.hituzi.co.jp/>

ISBN978-4-89476-705-8 C3080

造本には充分注意しておりますが、落丁・乱丁などがございましたら、
小社かお買上げ書店にておとりかえいたします。ご意見、ご感想など、
小社までお寄せ下されば幸いです。

謝辞

多数の研究仲間の援助のおかげで本書を書き上げることができた。彼らの助力に心より謝意を示したい。Svenja Adolphs、Karin Aijmer、Mark Davies、Costas Gabrielatos、Geoffrey Leech、Neil Millar、Richard Zhonghua Xiao の各氏には本書草稿の一部ないし全体を読んでもらい、有益な批評や示唆をいただいた(ただし、言うまでもなく最終版に対する責任は著者のみにある)。また、コーパス言語学と社会言語学の類似性に関して有益な議論を交わした Eivind Torgerson、クラスター分析に関連する様々な話題や網羅的な言語特性木の概念について長時間の議論を交わした Ghada Mohamed の両氏にも感謝申し上げたい(これらの議論はともに 5 章の記述の参考となった)。また、多すぎてここでは名前をあげられないが、他にも研究仲間や学生が適切な参考文献を教えてくれたり、出版前の書籍や研究論文の草稿に目を通す機会を与えてくれたりした。これらがなければ、本書でそうした研究に触れることはできなかったであろう。以上すべての人々に感謝申し上げる。また、Cambridge University Press の編集スタッフの高い専門性にも感謝したい。

日本語版への序文

著者は、ランカスター大学(Lancaster University)において、長年にわたり、日本から来た言語学専攻の大学院生の指導にあたっている。多くの学生がコーパス言語学に強い関心を抱き、やがて、その関心の種を携えて日本に帰国し、日本の大学で教育にあたっている。だが、日本のコーパス言語学に対するランカスター大学の関与はこれだけにとどまらない。というのも、ランカスターから帰国した多くの研究者が日本で英語学や英語教育の仕事に携わり、これによって、幅広い言語や言語事象に関心対象とする日本の言語学者が次第にコーパス言語学に引き寄せられていったからである。Tony McEneryは1998年に訪日し、東北大学の訪問教授として3か月間滞在した。コーパス言語学の教育を行っただけでなく、同大学の後藤斉氏の助力を得て、コーパス言語学の習得と活用を促進するべく、日本初となるオンラインの言語学研究資料データベースを立ち上げた。東北大学言語学研究室が管理するこのウェブサイトは、日本国内の研究者に対して、コーパス言語学の基礎理念と概念を紹介しようとしたものである[現在は対象を広げ、「国内言語学関連研究機関 WWW ページリスト」および「国内人文系研究機関 WWW ページリスト」として運用されている]。本書の読者の中にも、同ウェブサイトを通じて、コーパス言語学を初めて知った方が多数いることだろう。

日本でこのウェブサイトを立ち上げた頃から、コーパス言語学は非常に速いペースで発展してきた。ウェブサイトは今も有用なものであるが、著者はその後、コーパス言語学という話題について網羅的で最新の解説を行うべき時期が来たと感じ、本書の執筆に至った。本書で皆さんがお読みになるのはそうしたものである。中でも、コンコーダンスソフトウェアの発展といった話題は、1990年代にはそれほど大きな議論になるものではなかった。実際、使用できるソフトウェアパッケージはごくわずかであったし、その範囲も限定されていた。だが、上述のウェブサイトが使用できるようになった頃より、コンコーダンスシステムは速いペースで広範囲にわたって発展を見せている。以下でお読みいただけるように、本書においても、こうした検索技術が言語学の研究課題を扱う我々の能力に与えた影響について議論している。

現在では、コンコーダンスシステムの使用は、単に可能になったというだけでなく、もはや必須のものになったと言えよう。

コーパス言語学に関するウェブサイトや研究書を組み合わせれば、言語学者に対してコーパス言語学を教える目的はもはや達成されたと思われるかもしれないが、そう結論する前に、既存のコーパス言語学の文献の多くに大きな欠点が潜んでいることを考慮する必要がある。つまり、関連する資料の多くが英語で書かれているという問題である。これは英語話者である言語学者には好ましいことだが、ごく当然のこととして、英語話者以外の優秀な言語学者も多い。ゆえに、日本語で読めるコーパス言語学の入門書として、本書は日本のコーパス言語学にとって重要な一歩を示すものとなろう。本書を通して、英語以外の言語に関心を持っている幅広い言語学者が言語研究におけるコーパス研究手法に触れる機会を持ち、ひいては、英語コーパス学会(JAECS)をはじめとする各種の学術組織によって実践されている日本の卓越したコーパス言語学研究にさらなる刺激を与えることができればと願っている。

2013 年 10 月

Tony McEnery, Andrew Hardie

序文

本書のタイトル(原題)は *Corpus Linguistics: Method, Theory and Practice* である。タイトルが示唆するとおり、本書で扱う内容は、コーパス言語学が進展を遂げ、研究手法として使用されるに至る経緯、現代のコーパス言語学者が議論の対象としている主要な理論上の課題、言語学分野の内外においてコーパスを使用している研究者が実践上対処しなければならない問題などである。

一言で言えば、上記の内容を扱っている点にこそ、1990年代中葉から刊行され始めた多くのすぐれたコーパス言語学の入門書と本書の違いがあると考えたい。本書執筆の目的は、コーパス言語学の各種の分析手順の基本を紹介することではない。実際、本書では、コーパス調査の方法を分析手順ごとに記述することはしていない。コーパス分析の実例を紹介することはあるが、コンコーダンス、コロケーション、特徴語、また、コーパス検索ツールで得られるその他の一般的な出力結果を得る方法に関して詳細な解説を行うことはしない。こうした点はすでに他の本で包括的に扱われており(例: Biber *et al.* 1998; Hunston 2002; Adolphs 2006; McEnery *et al.* 2006; Hoffmann *et al.* 2008)、その説明をあえて繰り返す必要を感じなかったためである。

その代わりに本書が目的としたのは、コーパスを使用したこれまでの研究の中で重要だと思われる潮流を概観するとともに、コーパス使用の根幹をなす概念上の基本的問題点を紹介・説明し、場合によってはそれらに疑問を呈することである。本書の前半部では、研究分野としてのコーパス言語学を論じ、コーパス言語学者が関心を寄せる高次の実践的問題を考察する。つまり、均衡性や代表性の概念をはじめとするコーパス構築上の特性、コーパス分析ツールの開発と利用、著作権法や研究倫理に関わる問題、コーパスアノテーションの役割や限界、量的分析の役割などである。また、様々なコーパス言語学派や研究拠点における研究上の優先事項と主たる貢献についても概観する。これにより、各派・各拠点の方法論的な道具立てと、それがあてはまる場合には、彼らが目指した言語理論への貢献がともに検討されることとなる。一方、本書の後半部では、コーパス言語学そのものの境界を越え、言

語の多様性、言語変化、機能・認知言語理論、心理言語学の研究など、幅広いタイプの言語調査における現在のコーパス研究手法の役割と、当該研究でのコーパス使用で生じている問題について検討する。その上で、いわゆるコーパス言語学は長きにわたって孤立した存在であったが、現在では、コーパス言語学と、上述のような言語学の諸分野が大きく接近しつつあることを述べる。コーパス分析技術が特定領域の専門家のものとされる傾向はもはや薄れており、むしろ言語学全体に(さらには言語学以外の分野にまで)関係する重要な研究資源となっている。以上をふまえて主張したい点は、当該分野の将来のありようは、独立したコーパス言語学というよりも、言語学の中のコーパス研究手法といったものになるだろうということだ。

ただ、注記しておかないといけないが、こうした見方は必ずしも広く受け入れられているわけではない。とくに、新 Firth 派コーパス言語学の研究者の一部は、こうした見方に完全に反対している。この点は、6 章において当該の研究系譜を論じる際に触れる。万人の同意は期待できそうにない問題を論じる上で、著者には 2 つの選択肢があった。1 つ目は、著者自身の意見や理論・方法論上の好みで説明を恣意的に歪めることなく、中立を守って記述を進めてゆくことである。2 つ目は、著者自身の視点とそれに伴う偏りを率直かつ明示的に認めた上で記述を進め、できる限り自身の考え方を説明し、その正当性を主張しようとするものである。本書は 2 つ目の選択肢を取っている。コーパス研究における様々な系譜、言語学分野内外での幅広いコーパス使用、著者が望ましいと考えるコーパス言語学の将来像などに関する本書の議論は、結果的に、(Teubert 2005 の忘れがたい言い方に倣えば) McEnery & Hardie 版コーパス言語学とでも呼びうる立場の宣言文となっている。本書は中立性を装うことはしていないので、自分たち以外のコーパス言語学の研究系譜をめぐる議論において、実のところ、当該研究者が必ずしも同意しない形で、特定の研究学派を特徴付けている可能性もある。この最も顕著な例は、新 Firth 派研究者でない著者の立場が、新 Firth 派コーパス言語学をめぐる議論に大きく影響していることだ。ゆえに、この議論の全体像の理解に興味を持つ読者は、本書で言及する新 Firth 派の研究者自身が、本書と異なる前提・意見・方向性のもと、自らの視点で関連する問題を論じている文献をあわせて参照するよう薦めたい。本書は、適切と判断した場合には、各章末の文献案内において、こうした文献を列挙して紹介するようにした。たとえば、新 Firth 派理論についてであれば、本書 6 章および 8 章で行った議論の

別の面を明瞭かつ読みやすく紹介した本として、Tognini-Bonelli (2001) ないし Teubert & Čermáková (2004) を強く推奨したい。

文献案内では、適切な場合、原著論文と概観論文の両方を推薦している。また、各章末にはいくつか課題も用意した。これらは2種類に分かれている。本書が扱っている事柄や問題のいくつかについてさらに考察を深めたい読者には議論課題を薦める。加えて、実践的な演習課題も掲載しており、演習課題では、コーパス言語学の研究を実際にやってみる際に必要となる重要な手順のいくつかが練習できる。演習課題は、どれも、一定規模のコーパスと何らかのコーパス分析用ソフトウェアが使用できることを前提にしているが、特定のコーパスを用意する必要はない。自分のコンピュータ上に適当なコーパスがある場合は、いくつかあるコーパス分析ツールの中から任意のものをを用いて(ウェブ上で無償公開されているものもある)、コーパスデータを処理できる。コーパスデータを持っていない場合でも、ウェブ経由で使用・分析できる大規模な標準的コーパスが多数存在し、対象言語も複数に及ぶ。この場合、分析にはオンライン検索インタフェースが使用できる。こうしたインタフェースは一般公開されているものもあれば、無償で登録できるアカウント制を取っているものもある。オンライン検索インタフェースは、ここでは列挙できないほど数が多い(2章の議論を参照)。英語であれば、BNCweb、またはブリガムヤング大学(Brigham Young University)の開発したインタフェースを推奨する。ソフトウェアツールによって検索機能に違いがあるが、本書の演習課題はすべてのコーパス検索ツールで使用できる基礎的機能に限ったものとなっている。

今日、コーパス言語学の実践面ではウェブが中核的位置を占めるようになっており、本書でも、随所で各種のウェブサイトやウェブサービスを引証する必要があった。その際、本書では、ウェブアドレス情報の有用性が相対的に短期間に限られることを考慮して、本文中に記載するウェブアドレスを最小限にとどめるようにした[日本語版では原則として削除している]。それでも、時間の経過の中で記載したウェブアドレスの一部が変更されることは避けがたい。そこで、本書専用のサポートサイトを用意した。サポートサイト上には、必要に応じて更新されたリンク情報を掲載する。また、各章の課題の回答例など、他の補足的資料も掲載する。さらに、重要な点だが、単に数が多すぎたという理由で本書に記載できなかった追加の注記情報も数多く掲載する。ゆえに、各章を読了されたら、サポートサイトをご覧くださいこと

を勧めたい。サポートサイトのウェブアドレスは <http://corpora.lancs.ac.uk/clmtp> である。

訳者まえがき

「教科書は面白くない」というのは、およそどの分野でも定説であろう。ただ、書く側にも言い分はある。書き手自身の考えがあまり前に出ないよう、様々な見解を整理し、努めて中立的な立場から記述しようとすれば、面白さが入り込む余地はほとんど残されていないからである。訳者自身、コーパス言語学の教科書を書いてそういう思いを強くした。

したがって、本書刊行の知らせを聞いたときも、さして期待はしていなかった。これまでに少なからぬ数のコーパス言語学の教科書が書かれているが、細部を別にすれば、内容はどの本も基本的に類似していたからである。

ところが、実際に本書を入手し、最初の数章を読んだ瞬間に、自分の考えが根本的に間違っていたことに気付いた。本書の著者は、教科書の書き手に暗黙のうちに要請される中立性を自ら脱ぎ捨て(著者自身、「本書は中立性を装うことはしていない」とはっきり述べている)、自身の信念と視点から、言わば「McEnergy & Hardie 版コーパス言語学」とでも呼ぶべきものを書き上げていたからである。

とくに、新 Firth 派(コーパスそのものを独自の理論とみなす考え方)と、非新 Firth 派(コーパスを広義の研究手法とみなす考え方)をあえて対立させることで、コーパス研究の本質に鋭く切り込んでゆく著者の主張は独創的であり、その記述は刺激と面白さに満ちている。

一般的なコーパス言語学の教科書に慣れた目から見ると、著者の主張は時に過激に映るかもしれない。たとえば、下記の数行を引用するだけで、本書のユニークな性格は十分に伝わるであろう。

コーパス言語学にはいかなる将来の進展が予想されるのであろうか…論理的に考えて行きつく先は、独立した研究分野としてのコーパス言語学の消滅のように思われる。(9 章)

世界のコーパス研究を牽引する著者が、自身の研究分野の「消滅」を予言するのは、必ずしもコーパス研究の将来を悲観しているためではない。著者

は豊富な実例をあげ、コーパス研究が、談話分析・社会言語学・言語類型論・認知言語学・機能言語学・心理言語学といった言語学の諸分野はもちろん、文学・歴史学・脳科学・工学といった幅広い研究領域と統合されることで、さらに大きな貢献を果たす将来像を描き出し、そこに尽きるものないコーパスの可能性を見出している。この展望には説得力がある。

本書は、コーパス言語学の入門書のみならず、言語学全般の入門書としてもきわめてすぐれたものであり、多くの読者に読まれるべき本である。我が国において、コーパス言語学が、英語のみならずさまざまな言語の研究に影響力を広げつつある状況をふまえ、読了後、すぐさま日本語版の刊行を決意し、翻訳作業にあたってきた。この間、増補版の参考文献リストの提供を含め、訳者をサポートしてくれた2名の著者、また、出版事情の厳しい中、本書の価値に共感され、刊行を引き受けてくださったひつじ書房、同社編集部の森脇尊志氏に心より御礼申し上げる。

Brown Corpus 公開 50 周年の年に

石川 慎一郎

日本語版覚え書き

- 1) 言語学全般の入門書として読めるよう、専門用語および準専門用語はできる限り原語を併記し、必要に応じて訳注を付した。その際、原著で形容詞形で出ているものは名詞形に戻し、大文字表記も原則として小文字表記に統一した(例: emergentist → emergentism、Conceptual Metaphor Theory → conceptual metaphor theory)。
- 2) 原注(本文注・巻末注)は丸括弧で、訳注は角括弧で表した。
- 3) 原著の巻末注については、重要と判断されるものを精選し、本文内に原注として記載した。なお、将来変更が予想されることから、ウェブアドレスに関する注記は原則として省略した。
- 4) スペルアウト語と頭字語の併記については、一部を除き、原則としてスペルアウト語を先に記載した。また、原著に反復して使用されている頭字語のうち、一般化していないと思われるものについては(例: ECL、CMT など)、該当する日本語に置き換えて記載した。
- 5) 文中での先行研究の列挙はセミコロンで区切りを示した。著者が複数の場合の表記は A & B とした。
- 6) 原著における強調のための引用符や斜字体は、それがないと意味理解が難しくなる場合を除き、原則として削除した。
- 7) 原著の各章末(最終章除く)には、節番号のない Further Reading、Practical Activities、Questions for Discussion という3つのコラムが掲載されている。日本語版ではこれら3つを「発展学習のために」という新設のセクションにまとめ、それぞれ「文献案内」「演習課題」「議論課題」として新規に節番号を与えた。

目 次

謝辞	ix
日本語版への序文	xi
序文	xiii
訳者まえがき	xvii
日本語版覚え書き	xix

第 1 章 コーパス言語学とはなにか	1
1.1 はじめに	1
1.2 言語様式	5
1.3 コーパス検証型言語学とコーパス駆動型言語学	8
1.4 データ収集法	9
1.4.1 モニターコーパス	9
1.4.2 コーパスとしてのウェブ	10
1.4.3 標本コーパス型データ収集法	12
1.4.4 均衡性、代表性、比較可能性	15
1.4.5 簡易コーパスと少数言語・危機言語	16
1.5 アノテーションコーパスと非アノテーションコーパス	19
1.6 網羅的データ説明原則と部分的データ選択	21
1.6.1 網羅的データ説明原則、反証可能性、再現可能性	21
1.6.2 部分的データ選択が(必ずしも)問題ではない場合	24
1.7 単一言語コーパスと複数言語コーパス	26
1.8 まとめ	30
1.9 発展学習のために	31
1.9.1 文献案内	31
1.9.2 演習課題	32
1.9.3 議論課題	34

第2章 コーパスデータの利用と分析	37
2.1 はじめに	37
2.2 コーパスは言語学のすべてのリサーチクエスションの解となるか	40
2.3 コーパスのアノテーション	43
2.3.1 メタデータ、マークアップ、アノテーション	43
2.3.2 アノテーションの一貫性	47
2.4 コンコーダンスの表示	52
2.5 コーパス分析ツールの史的概観	55
2.5.1 第1世代コンコーダンス	56
2.5.2 第2世代コンコーダンス	58
2.5.3 第3世代コンコーダンス	60
2.5.4 第4世代コンコーダンス	65
2.6 コーパス言語学における統計	73
2.6.1 記述統計	74
2.6.2 記述統計を超えて	77
2.7 まとめ	81
2.8 発展学習のために	82
2.8.1 文献案内	82
2.8.2 演習課題	83
2.8.3 議論課題	84
 第3章 ウェブの使用における法と倫理	 85
3.1 はじめに	85
3.2 ウェブと法的問題	85
3.3 倫理的問題	90
3.3.1 調査回答者と倫理	91
3.3.2 コーパス開発者と倫理	96
3.3.3 コーパス配布者と倫理	98
3.3.4 コーパス使用者と倫理	99
3.3.5 倫理的問題を含む研究	101
3.4 まとめ	104

3.5 発展学習のために	104
3.5.1 文献案内	104
3.5.2 演習課題	105
3.5.3 議論課題	105
第4章 英語コーパス言語学	107
4.1 はじめに	107
4.2 ユニヴァーシティ・カレッジ・ロンドン	111
4.3 ランカスター大学	114
4.4 バーミンガム大学	119
4.5 ルーヴァン＝カトリック大学	123
4.6 ノッティンガム大学	127
4.7 北アリゾナ大学とアメリカ	133
4.8 まとめ	136
4.9 発展学習のために	137
4.9.1 文献案内	137
4.9.2 演習課題	138
4.9.3 議論課題	139
第5章 共時的・通時的多様性に対するコーパス準拠型研究	141
5.1 はじめに	141
5.2 古英語から近代英語に至る通時的变化	142
5.2.1 汎用型歴史的コーパスの代表例	142
5.2.2 特殊な歴史的コーパス	144
5.3 現代英語における通時的多様性	145
5.3.1 Brown Family	145
5.3.2 Brown Family の分析から得られた知見	150
5.4 多様性研究のための多次元分析法(MD 法)	156
5.4.1 多次元分析法の概観	156
5.4.2 多次元分析法の応用	162

5.4.3 Biber の多次元分析法への批判	167
5.4.4 多次元分析法：まとめ	172
5.5 コーパスと言語変異論系社会言語学	172
5.6 まとめ	176
5.7 発展学習のために	177
5.7.1 文献案内	177
5.7.2 演習課題	177
5.7.3 議論課題	180
 第 6 章 新 Firth 派コーパス言語学	 181
6.1 はじめに	181
6.2 コロケーション	182
6.3 談話	197
6.4 意味的韻律と意味的選好性	201
6.5 語彙と文法	211
6.5.1 イディオム原則	211
6.5.2 パタン文法	213
6.5.3 語彙ブライミング	215
6.6 理論としてのコーパス派と手法としてのコーパス派	218
6.6.1 データと理論	218
6.6.2 テキスト標本抽出	226
6.6.3 コーパスアノテーション	228
6.6.4 コーパス言語学と既存理論	233
6.7 まとめ：コーパス言語学に対する Sinclair の貢献	242
6.8 発展学習のために	244
6.8.1 文献案内	244
6.8.2 演習課題	244
6.8.3 議論課題	246