

英語英文学研究と コンピュータ

齊藤俊雄 編

英潮社
EICHOSHA

英語英文学研究と コンピュータ

齊 藤 俊 雄 編



英 潮 社

英語英文学研究とコンピュータ

1992年 3月30日 初版発行 検印廃止 定価5,800円(本体5,631円)

編 者 齊 藤 俊 雄

発 行 者 土 岐 省 三

発 行 所 髙 英 潮 社

東京都千代田区神田神保町2-28日下ビル(〒101)

電話 東京 (3263) 1641(代)・6171 番

振 替 口 座 ・ 東京 7-55407 番

〔電植・株式会社キャップス〕

ISBN4-268-00034-8 C3082 P5800E

は し が き

本書はコンピュータ利用による英語学・英文学研究の論文集である。本論文集が編纂されるに至った動機は、近代英語協会第7回大会（1990年5月就実女子大学において開催）におけるシンポジウム「近代英語研究とコンピュータ」にある。筆者の司会のもとに中村純作徳島大学教授の「Brown Corpus等のコーパス利用による英語分析」、今井光規大阪大学教授の「パソコンを使って用例カードを作る」、久屋孝夫西南学院大学助教授の「パソコンを利用した初期近代英語の基礎的研究」の研究発表があり、当初の予想を遙かに上回る出席者を得て、熱のこもった討議が行われた。近代英語の学会であるにもかかわらず、聴衆の中には中世英語英文学者たちの顔も見られ、コンピュータ利用に対する英語英文学者の期待に満ちた熱い眼差しを感じざるを得なかった。また同学会の懇親会でも、コンピュータの利用を模索している人たちから様々な熱心な質疑を受けた。

このようなコンピュータ利用の趨勢は、ここ4、5年前から徐々に勢いを得てきて、日本英語学会第6回全国大会（1988年11月青山学院において開催）でも水鳥喜喬教授の司会で「コンピュータ・コンコーダンスで何が出来るか—OE・ME研究を中心に—」というシンポジウムが行われている。さらに学会発表もコンピュータ利用のものが見られるようになってきている。

しかしながら、筆者は近代英語協会の懇親会の席で多くの研究者と意見を交換して、英語英文学者の多くが、コンピュータ利用の英語学・英文学研究に関心があっても、まだいかに取り組んでいったらよいのか、その方法に戸惑っているように感じられたのである。従って、今必要とするものは、コンピュータ利用の研究の手引きになるようなもの、研究のモデルになるような書物であると確信するに至ったわけである。幸いにして、近代英語協会の熱心な賛助者である英潮社社長土岐省三氏の熱心な勧めもあって、非力も省みず本論文集の編纂に踏み切った次第である。

編集方針として、本書がわが国におけるコンピュータ利用による英語学・英文学研究の現在の水準を示すと同時に、「コンピュータ利用英語学・英文学研究の事始め」といったものになるように、コンピュータ利用による研究の基本的情報を与えるよう心掛けた。

具体的には、学会誌等の優れた既発表論文を集めると共に、現在コンピュータ利用を活発に行っている研究者に書き下ろし論文を依頼し、さらに筆者が基本的情報を与える序説を執筆することにした。その結果集まった論文は、時代的には古英語時代から現代にまで及び、分野的には語学研究もあれば、文学研究もあり、またコンピュータ利用の点で非常にソフィスティケートなものがあれば、また非常に素朴なものもある。さらに内容的には序論的・入門的なものもあれば、また非常に専門的なものもあるといった具合である。

読者はこれらの論文を読むことによって、各自のコンピュータ・リテラシーの度合いに応じて様々な刺激を受けることであろう。論文執筆者の一人が筆者への私信で日本のコンピュータ利用英語学・英文学研究は世界から10年は遅れていると述べているが、本論文集が日本におけるコンピュータ利用の英語学・英文学研究の前進への一つの刺激剤になれば、その目的を達したことになる。

ここで本論文集編集のきっかけを与えて下さった近代英語協会の荒木一雄会長と天野政千代事務局長に心から感謝するとともに、既発表5論文の転載を許可して頂いた関係諸機関に謝意を表する次第である。また本書の校正に助力をして頂いた大阪大学言語文化部長室の徳野ほづみさんにも感謝の意を表したい。

最後になったが、この論文集の出版に当たって並々ならぬご援助を賜った英潮社社長の土岐省三氏、並びに表と図版の多い面倒な本書の編集を担当して頂いた中條方章氏に衷心からお礼を申し上げる。

1991年3月

齊 藤 俊 雄

謝 辞

既刊論文の本書への転載を許可して頂いた次の諸機関に感謝する。

日本中世英語英文学会

近代英語協会

西南学院大学学術研究所

寺澤芳雄教授還暦記念論文集編纂委員会

編 者

目 次

はしがき

コンピュータによる英語英文学研究序説 …………… 齊藤俊雄 1

第一部 近代英語英文学研究

ブラウンコーパス等のコーパス利用による

英語テキスト分析 …………… 中村純作 21

スタインベックの小説

——コンピュータによる文体分析の可能性——

………… スコット・ピュー 52

特定作家の文体研究

——D. H. ローレンスの場合——………… 西村道信 72

『序曲』を構成する語彙について …………… 中川 憲 93

パソコンによる自家製コンコーダンスと

聖書英語の研究 …………… 橋本 功 114

Authorship とコンピュータと脚韻語

——16世紀前半の劇の言語について——………… 久屋孝夫 130

第二部 中英語英文学研究

『カンタベリ物語』における shall と will の出現率と

文体の関連について …………… 西出公之・川端 喬 165

パーソナル・コンピュータによる *Sir Thopas* の言語の

計量的研究 …………… 中尾佳行・松尾雅嗣 177

The House of Fame における対照語法とその用法

——特に 'soth' と 'fals' について——………… 地村彰之 195

中英語韻文ロマンスにおける Gan の機能

——コンピュータ編纂のコンコーダンスを利用して——

……水谷洋一 214

Octovian Imperator, Sir Launfal, Libeaus Desconus

における脚韻語句 ……………西村秀夫 229

中英語韻文ロマンスのコンコーダンス編纂計画と

その現状 ……………齊藤俊雄・今井光規 249

第三部 古英語英文学研究

Ælfric と Wulfstan

——Toronto Corpus の計算機処理による語彙の比較——

……山縣宏光 275

レマタイゼーションそれ以降：ウルフスタンの場合 ……鈴木重樹 290

OED に見られる古英語頭韻詩 *BEOWULF* の引用

——OED 及び CD-ROM 版 OED を検討する——

……………渡辺秀樹・岩根 久 304

付 録

機械可読テキスト・リスト ……………333

コンピュータによる英語英文学研究序説

齊藤 俊雄

1. はじめに

最近のコンピュータ、特にパソコンの発達・普及によってコンピュータが様々な分野に利用されている。現代社会はコンピュータ無しでは考えられないのが現状である。我々が研究している英語学・英文学の分野でも例外ではありえない。わが国の英語学・英文学界では、特に英語学界では理論言語学のめざましい発展のみが目につきがちであるが、目を海外に転ずると、コンピュータ利用の *corpus linguistics* の発展とその成果は目を見張らせるものがある。

このような状況から、わが国でも徐々にコンピュータ利用の英語学・英文学の研究が興ってきている。1988年の日本英語学会第5回大会において水鳥喜喬教授司会でシンポジウム「コンピュータ・コンコーダンスで何ができるか——OE・ME研究を中心に——」が行われ、1990年の近代英語協会第7回大会では筆者の司会でシンポジウム「近代英語研究とコンピュータ」が行われ、予想を遙かに超える多数の出席者を得た。また諸学会におけるコンピュータ利用の英語学・英文学研究成果の発表も目につくようになってきた。

現在コンピュータ利用の研究を始めようと考えている英語学者、英文学者も多いと考えられる。本書は、はしがきにある通り、「コンピュータ利用英語学・英文学研究の事始め」といったものであるが、ことにこの章は、そのようなこれからコンピュータを利用しようと考えている人々に、極めて基礎的な情報——主として文献——を与えようとするものである。しかも筆者の研究範囲の狭さから、極めて限られた分野の情報になることを恐れるものである。

2. 基本的文献

2. 1. 入門書・概説書

コンピュータ利用の英語学・英文学研究を始めるに当たって、最初に必要なものは入門書・概説書の類であろうが、英語学・英文学のみを対象にしたものはあまり見られない。広く人文科学全般、あるいは言語、文学を対象にした英語で書かれたものに、次のようなものがある（刊行順）。

1. Susan Hockey, *A Guide to Computer Applications in the Humanities*, Duckworth, London; Johns Hopkins University Press, Baltimore, 1980.
2. Robert Oakman, *Computer Methods for Literary Research*, The University of Georgia Press, 1984.
3. Christopher Butler, *Computers in Linguistics*, Blackwell, 1985.
4. Derek Andrews and Michael Greenhalgh, *Computing for Non-scientific Applications*, Leicester University Press, 1987.
5. Paula R. Feldman and Buford Norman, *The Wordworthy Computer: Classroom and Research Applications in Language and Literature*, Random House, New York, 1987.
6. B. H. Rudall and T. N. Corns, *Computers and Literature: A Practical Guide*, Abacus Press, 1987.
7. Sebastian Rahtz (ed.), *Information Technology in the Humanities: Tools, Techniques and Applications*, Ellis Horwood, Chichester, 1987.
8. George W. Smith, *Computers and Human Language*, Oxford University Press, 1991.

それぞれに書誌があり有益である。3には語学・文学研究の様々な分野におけるこれまでの研究成果の概括があり、たいへん便利である。

日本語で書かれたものに、次のようなものがある。

1. 長瀬真理・西村弘之『コンピュータによる文章解析入門——OCPへの招待——』オーム社, 1986.
2. 竹蓋幸生『英語教師のパソコン』エデュカ, 1986.

3. ———『英語教師のパソコンII』エデュカ, 1987.

4. 村山皓司編『異文化を知るための情報リテラシー——外国語と外国文化研究教育におけるコンピュータ利用入門』法律文化社, 1990.

1はOCP (Oxford Concordance Program) の入門書であるが、文体統計学 (stylostatistics) の章があり、2, 3では英語(教育)研究に必要な様々なプログラムが紹介され、そのプログラムのフロッピーが入手できるようになっている。4は外国語と異文化の研究におけるコンピュータ利用入門のみならず、プログラミング入門までと範囲が広い。

2. 2. 論文集

コンピュータ利用の語学・文学研究の論集も参考になる。学会発表の論文を集めたいわゆる 'proceedings' が多く、様々な分野でのコンピュータ利用の成果を知ることができる。なお専らコーパス言語学の論集は、コーパスの項を参照のこと。

1. R. A. Wisbey (ed.), *The Computer in Literary and Linguistic Research: Papers from a Cambridge Symposium*, Cambridge University Press, 1971.

2. A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith (eds.), *The Computer and Literary Studies*, Edinburgh University Press, 1973.

3. J. L. Mitchell (ed.), *Computers in the Humanities*, Edinburgh University Press, 1974.

4. Alan Jones and R. F. Churchhouse (eds.), *The Computer in Literary and Linguistic Studies*, University of Wales Press, Cardiff, 1976.

5. S. Lusignan and J. S. North (eds.), *Computing in the Humanities*, University of Waterloo Press, Ontario, 1977.

6. D. E. Ager, F. E. Knowles, and J. Smith (eds.), *Advances in Computer-aided Literary and Linguistic Research*, The University of Aston in Birmingham, 1979.

7. R. W. Bailey (ed.), *Computing in the Humanities*, Ann Arbor, Michigan, 1981.

8. P. C. Patton and R. A. Holoien (eds.), *Computing in the Human-*

- ities, Heath, Lexington, Mass., 1981.
9. L. Cignoni and C. Peters (eds.), *Computers in Literary and Linguistic Research*, Pisa, 1982.
 10. J. Hamesse and A. Zampolli (eds.), *Computers in Literary and Linguistic Computing*, Champion-Slatkine, Paris, 1985.
 11. K. C. Cameron, W. S. Dodd, and S. P. Q. Rahtz (eds.), *Computers and Modern Language Studies*, Ellis Horwood, Chichester, 1986.
 12. David S. Miall (ed.), *Humanities and the Computer—New Directions*, Clarendon Press, Oxford, 1990.

2. 3. 雑誌・年鑑

新しい研究は、どうしても雑誌の論文に頼ることになるが、コンピュータ利用による人文系学問研究や語学・文学研究のための国際的な雑誌に、次の2誌がある。

1. *Computers and the Humanities* (Kluwer Academic Publishers, the Netherlands, 1966-).
 2. *Literary & Linguistic Computing* (Oxford University Press, 1986-).
- 両誌とも季刊である。それぞれこの方面の代表的な学会の機関誌である。1はThe Association for Computers and the Humanities (ACH)の機関誌で最も歴史の古いもので、これまで発行所が転々と変わってきたが、語学・文学のみならず、広く人文系のコンピュータ利用研究を扱う。2はThe Association for Literary and Linguistic Computing (ALLC)の機関誌であり、*ALLC Journal* (1973-85), *ALLC Bulletin* (1980-85)の後身で比較的新しいが、語学・文学の専門誌であり、もちろん英語学・英文学以外も扱う。両誌ともコンピュータ利用の語学・文学研究の諸学会の情報を得るのに都合がよい。

英語専門の雑誌に次のものがある。

1. *ICAME Journal* (Norwegian Computing Centre for the Humanities, Bergen).
- これはThe International Computer Archive of Modern English (ICAME)の機関誌であり、No. 10 (1986)まで*ICAME NEWS*と呼ばれていたもので、年刊である。このICAMEというアーカイヴは、英語関

系の機械可読コーパス (machine-readable corpora) のクリアリング・ハウスの役割を果たしているところであり、その機関誌であるこの雑誌は、世界における英語関係のコーパス編纂の進捗状況やコーパス言語学の研究成果を知らせてくれる、英語研究者にとってまことに便利なものである。

年鑑には次のものがある。

1. Ian Lancashire and Willard McCarthy (eds.), *The Humanities Computing Yearbook 1988*, Clarendon Press, Oxford, 1988.
2. Susan Hockey and Nancy Ide (eds.), *Research in Humanities Computing 1989/90*, Clarendon Press, Oxford, 1991

1 は人文系のコンピュータ利用研究についての画期的な年鑑である。書名が少し誤解を招きかねないが、1 年だけの業績を集めたものでなく、人文系学問へのコンピュータ利用の初期の段階から現在までの、人文系のすべての分野における研究業績、ソフトウェアを網羅的に紹介している。研究者にとっては必須の参考文献であろう (最近 Lancashire の単独編集で 1989/90 年度版が出たようであるが、筆者未見)。

2 は新しく毎年刊行されることになった年鑑で、筆者は校正の時点では未見であるが、ACH/ALLC の年次大会の研究発表を中心にしたもので、人文系のコンピュータ援用研究の動向を知るのに欠かせない年鑑になるものと思われる。

3. コーパス

コンピュータで英語学・英文学を研究するには、コンピュータで読める資料がまず必要なことは言うまでもない。このような機械可読テキストの集積をコーパス (corpus) という。世界で最初の大規模な英語コーパスは、米国の Brown University によって編集された The Standard Corpus of Present-Day Edited American English (通称 Brown Corpus) である。このコーパスに基づいて米語のコンピュータによる研究が行われたが、これが刺激となって、様々なコーパスが各地で編集され、それを利用した研究が盛んになり、いわゆるコーパス言語学 (corpus linguistics) という研究分野が生まれた。この項では代表的なコーパスを紹介する。

3. 1. The Standard Corpus of Present-Day Edited American English

この通称 Brown Corpus は、1961年に米国で出版された15のジャンルを代表する各2000語の500サンプルよりなる約100万語のコーパスであり、1963-64年に編集された。ジャンルの内容については本書の中村論文に詳しい。その後文法標識を付したコーパス (grammatically-tagged version) も作られた。

このコーパスによる米語研究の代表的な成果に次のものがある。

1. Henry Kučera and W. Nelson Francis, *Computational Analysis of Present-Day American English*, Brown University Press, Providence, 1967.
2. W. Nelson Francis and Henry Kučera, *Frequency Analysis of English Usage : Lexicon and Grammar*, Houghton Mifflin, Boston, 1982.

1 はオリジナル版の分析報告であり、2 は tagged version の分析結果である。

3. 2. The Lancaster-Oslo/Bergen Corpus of British English (LOB Corpus)

これは Brown Corpus の英国英語版とも言うべきものである。英国英語と米語の比較研究を念頭に置いて編集されたもので、Brown Corpus と同様に、1961年に英国で出版された15のジャンルを代表する各2000語の500のサンプルの約100万語よりなる。Lancaster 大学によって始められ、Oslo 大学と Norwegian Computing Centre for the Humanities, Bergen の協力で完成し、1978年に公開されている。文法標識を付した版も完成している。Lancaster 大学では自動的文法標識付加プログラムの開発が進み、Roger Garside, G. Leech and G. Sampson (eds.), *The Computational Analysis of English—A Corpus-Based Approach* (1987) によれば、CLAWS と呼ばれる tagging program での成功率は、96~97%の高さを示している。

このコーパスによる重要な研究成果に次のものがある。

1. K. Hofland and S. Johansson, *Word Frequencies in British and American English*, Norwegian Computing Centre for the Humanities, Bergen, 1982.

2. S. Johansson and K. Hofland, *Frequency Analysis of English Vocabulary and Grammar Based on the LOB Corpus*, 2 vols, Clarendon Press, Oxford, 1988-89.

1はオリジナル版の分析結果で、Brown Corpusとの比較がなされており、興味深い。2は tagged version に基づいた、コンピュータによる自然言語の品詞分類の貴重な成果であり、将来のコンピュータによる品詞研究、統語法研究への道を開くものである。

3. 3. London-Lund Corpus of Spoken English

これは University College London の Randolph Quirk 主宰の the Survey of English Usage の一環として始められ、Lund 大学の Jan Svartvik らによって the Survey of Spoken English 計画として1979年完成されたものである。当初英国の教育ある話手の英語の約5000語のテキスト87 (約435,000語) よりなっていたが、最近13テキストが追加され、100テキスト約50万語のコーパスになった。自然な私的な会話から演説、説教、講義までである。そのコーパスの一部が刊行されている。

Jan Svartvik and Randolph Quirk (eds.), *A Corpus of English Conversation*, Lund Studies in English 56, C. W. K. Gleerup, Lund, 1980.

また J. Svartvik (ed.), *The London-Lund Corpus: Description and Research*, Lund University Press (1990) によってこのコーパスの詳細な内容と研究成果 (200点以上のリスト) を知ることができる。

なお、付言すると、有名な the Survey of English Usage (SEU) はコンピュータ援用以前の現代英語コーパス、データベースであるが、その蓄積された貴重な資料は、現在 Sidney Greenbaum 教授の指導のもとに IBM UK Scientific Centre の Geoffrey Kaye の協力を得て機械可読のデータベース化が行われている¹⁾。

以上の3コーパスと the Survey of English Usage のコーパスの構成内容を知るには、N. Oostdijk, "A Corpus Linguistic Approach to Linguistic Variation"²⁾ が便利である。

3. 4. The Birmingham Collection of English Text (Birmingham Corpus)

J. M. Sinclair を中心とした Birmingham 大学英文科によって編集され

たものである。主に1960-81年の作品の約2000万語に及ぶ一般英語のテキストコーパスで、うち 'Main Corpus' は約730万語、'Reserve Corpus' は約1300万語である。

元来 COBUILD Lexicographic Project としてスタートし、*Collins COBUILD English Language Dictionary* (1987) の出版によってひとまず当初の目的を遂げたもの。現在このグループはさらに辞書類の編纂と英語の collocation の研究などを推進している。J. M. Sinclair の近著に *Corpus, Concordance, Collocation*, Oxford University Press (1991) がある。

3. 5. The Toronto Corpus of Old English in Computer-Readable Form

これまでに取り上げたコーパスは、すべて現代標準英語のコーパスであるが、これは古英語のコーパスである。Toronto 大学において現在進められている古英語辞書編纂のために古英語の資料全部を機械可読化したもので、300万語にのぼる。現在古英語の研究はこのコーパスなしでは考えられないが、その内容については、本書の山縣論文に詳しい。このコーパスを元にして次のコンコーダンスが編纂されている。

1. R. L. Venezky and A. P. Healey (eds.), *A Microfiche Concordance to Old English*, Newark and Toronto, 1980.
2. R. L. Venezky and S. Butler (eds.), *A Microfiche Concordance to Old English: High Frequency Words*, Newark and Toronto, 1985.

3. 6. The Helsinki Corpus of English Texts: Diachronic and Dialectal

このコーパスは通時的資料と方言資料のコーパスである。これは1984年以来ヘルシンキ大学英文科が Matti Rissanen 教授主宰で編纂中のもので、約150万語にのぼる8世紀から18世紀にわたる通時的資料の部分 (Diachronic Corpus) と伝統的な土地言葉を話す1970年代の年輩者に対するインタビューの録音に基づいた約40万語の方言資料の部分 (Dialectal Corpus) からなる。このコーパスは、通時的資料の部分が最近完成して公開されたばかりであって、現代英語のコーパスほどよく知られていないが、英語史的な研究に極めて重要な役割を果たす可能性のあるコーパスであり、少し詳しく紹介する。このコーパスの通時的コーパスは basic part と supplementary part よりなり、前者は次の表の通りである³⁾。

EModE. BRITISH

<i>Words</i>				<i>Words</i>			
I	1150-1250	108,000	18.1%	I	1500-1570	179,300	33.9%
II	1250-1350	96,400	16.2	II	1570-1640	189,200	35.7
III	1350-1420	181,400	30.4	III	1640-1710	161,000	30.4
IV	1420-1500	211,100	35.3				
		596,900	100.0			529,500	100.0

なお supplementary part は, Scots (現在300,000語) と early American English (現在155,000語) よりなる。Old English の部分は前述の Toronto Corpus より選んだものである。

通時的な言語変化は共時的な変異 (variation) を通じて解明すべきであるという考えのもとに、社会史的、社会言語学的な面から各時代の様々なタイプのテキストが選択されている。Merja Kytö and Matti Rissanen (1988) によれば⁴⁾、テキストのジャンル (text prototype と言っている) は 1. stipulation, 2. science, 3. secular instruction, 4. religious instruction, 5. non-fictive narration, 6. fictive narration, 7. correspondence, 8. drama, 9. proceeding の 9 つよりなり、さらに細かく下位区分される。

この Helsinki Corpus は、既にいくつかのパイロットスタディが行われ、その有効性が立証されている⁵⁾。

3. 7. The Century of Prose Corpus

さらに 'historical corpus' には Louis T. Milić 教授の編集する The