

英語コーパス言語学の歴史的背景

吉 野 貴 好

Historical Background of English Corpus Linguistics

Kiyoshi YOSHINO

Summary

With the development and the rapid spread of computer and Internet in the last several years, two fields of English studies have lately attracted considerable attention. One is CALL (Computer Assisted Language Learning), and the other is English Corpus Linguistics. I deal with the latter in this paper.

Why has it attracted considerable attention, then? It is because that computer corpus is extremely effective in a field of Corpus Linguistics and English Education such as a study of vocabulary, dictionary editing and corpus of learner 's English, etc. as a tool. To take a simple example, we can very easily know by utilizing English corpus that the expression ' make it a rule to do ' is not a mistake, but it is an extremely few to use it in written and spoken language. Corpus is principally coming into the limelight for high validity currently.

English Corpus Linguistics is unfamiliar even to an English teacher. So our concern is to consider what it is. The method to answer this question is to reflect the historical background of English Corpus Linguistics. A past is not a target put before us, but it is a target to learn. I hope the study of its history may contribute to a human language study.

はじめに

近年の急速なコンピュータ及びインターネットの発達と普及に伴って、現在、英語教育や英語学の分野で二つの領域が大きな注目を集めている。一つはCALL(Computer Assisted Language Learning)であり、他の一つは英語コーパス言語学(English Corpus Linguistics)である。拙稿では後者を扱う。何故注目を集めているかということ、電子コーパスが、語彙研究、文法研究、英語史研究、文体論研究、辞書編纂、電子辞書編纂、語学教育、学習者コーパス等々のコーパス言語学(Corpus linguistics)や英語教育の分野で極めて有効なツールであることが実証されてきたからである。例えば、鷹家秀史氏と須賀廣氏の共著『実践コーパス言語学』より、語法に関する実践の一例として‘make it a rule to do’を挙げて、コーパスの有用性を証示してみたい。

この表現は数多くの大学で頻繁に入試問題として出題されているし、受験問題集にも頻出する。受験生がよく用いている辞書の一つ、『小学館 プログレッシブ英和中辞典』には次のように説明されている。「...することにしている、...するのを常としている：I make it a ~ never to borrow money. 金は借りないことにしている。文語的なので口語では、generally, usually, as a rule を用いるのが普通。」恥ずかしさを禁じえないが、筆者自身この表現を必死で覚えて、英作文や日常会話で使ったことが多々ある。では、この表現は現実の英語圏世界で頻繁に用いられる表現なのだろうか？少なくとも文語ではよく用いられるのだろうか？こうした疑問に英語コーパス(English corpus)は明瞭に答えてくれる。つまり、‘make it a rule to do’は、5千万語のコーパスが利用できるCOBILDDirectのなかに次のような一例があるだけである。“... for I had told him that I made it a rule to dispense with the evening...” また、アメリカ話し言葉コーパスにも、TIMEや世界最初のコーパスであるBrown Corpusなどから作成した300万語のアメリカ書き言葉コーパスにも、一例もない。さらにアメリカ英語のBrown Corpusと並んで著名なイギリス英語のLOBコーパス(100万語)にはたった一例あるだけである。⁽¹⁾以上のことから、‘make it a rule to do’という表現は、チョムスキー流に言えば正しいにしても、現実には、口語でも文語でも、まず使用されることのない表現であることがわかる。有用性の視点からはナンセンスな表現といえるだろう。

また、ある高名なヘーゲル学者は、ヘーゲルの著作をe-textの形で収めたCD-ROMが発表されたとき、自分の数十年の労苦は一体何だったのか、と嘆いたそうである。つまり彼は、膨大な時間をかけヘーゲル作品のある語の用法やコロケーションをしらべ、それに基づいてヘーゲルの哲学思想を研究してきた訳であるが、今ではWordsmithやTxtana等々のシェアウェアソフトの機能を用いて(つまり、単語頻度リストやKWIC表示とコンコーダンサー等々)短時間で調べることができる。彼は失った時間の大きさを嘆いたのである。

このように、コーパスは一義的にはその有用性のために現在脚光を浴びている訳である。そこで、英語教員に耳新しいこの「コーパス」とか「英語コーパス言語学」が如何なるものであるかを考察

して行こう。その手法としては歴史的に反省して行いたい。過去は単に無意味に我々の前に置かれてある対象ではなく、学ぶ対象としてある。英語コーパス言語学の歴史を学ぶことによって、理論的にも実践的にも今後のその方向性や在るべき姿も見えてくるかもしれないし、また人間の言語研究に何ほどか資するかもしれない。

一 英語学と英語コーパス言語学

まず、(チョムスキーやコンピュータ・コーパス以前の)伝統的英語学の内容を概観してみたい。人間言語の科学的研究が言語学(linguistics)であるが、英語という特定の言語の科学的研究が英語学(English linguistics)である。その英語学の領域は、大まかに分類すれば、記述英語学(English descriptive linguistics)、歴史英語学(English historical linguistics)、そして地理英語学(English geolinguistics)とに分類される。記述英語学は共時英語学(English synchronic linguistics)とも呼ばれ、主として音声・音素を扱う。歴史英語学は通時英語学(English diachronic linguistics)とも呼ばれ、主として英語の形態論、統語論を扱う。地理英語学は、世界各地の英語の分布、政治・経済・社会・文化等々における各国語間の相互作用といった英語の実用面を扱う。英語学を記述英語学や歴史英語学にとらわれず分類する方法として、英語音韻論(English phonology)と英語文法論(English grammar)に大別する分類法がある。ここで更に英語音韻論は、英語音声学(English phonetics)と英語音素論(English phonemics)に分類され、英語文法論は、英語形態論(English morphology)と英語統語論(English syntax)に分類される。この他、英語学には英語意味論(English semantics)と英語文字論(English graphemics)とがある。従来英語音韻論と英語文法論とが英語学の主要分野と見なされてきた。では、英語学の特質とは何だろうか?英語学は主として1930年代後半に発達してきたが、それは、科学的研究を特質とする。つまり、客観的な実験、観察に基づく経験的方法をとるということである。主観的態度を極力避ける。なぜなら、それは対象を相対化する可能性があり、精密性を欠く可能性があるからである。⁽²⁾以上が伝統的英語学観であるが、当然現在でもこの見方に基づいて大学での講義やゼミ等々も行われている。しかし、チョムスキー(Noam Chomsky,1928 -)が現れて以降、経験に拠らず言語使用の創造性に着目する内面の言語学の研究も進捗し、ここ40年ほど彼の合理主義的言語論は英語学の主流領域を獲得している。

ところで、近年技術革新の進展や量産化技術の確立等々により機能向上やコスト低下が齎され、コンピュータが急速に普及してきた。特に、主としてコンピュータを端末にしたインターネットの広がりには目を見張るものがある。目下、欧米だけでなくアジア・日本においてもコンピュータ・インターネットを社会的インフラとしてネットワーク社会が構築されつつある。こうした社会内部の変容は、英語学の分野にも影響を及ぼしている。つまり、コンピュータの大きな特性として大規模データの保存可能機能が挙げられるが、この機能を用いて従来のコーパスを電子化して利用する試みが言語、特に英語において生じてきた。即ち、従来の対象である自然言語を自然言語で処理

する言語学というよりも、人工言語をツールとして自然言語を処理する言語学、つまり、英語コーパス言語学の胎頭である。それは応用言語学(applied linguistics)の一形態ともいえるかと思う。データをCD-ROMの形で保存するパッケージソフトの発達がコーパス言語学者の研究の負担を和らげてきたし、取り分け、1995年以降の急速なインターネットの社会への浸透とハードディスクの大容量化の実現により、従来手間暇をかけスキャナで取り込んでいた大量のデータが、ネット上からダウンロードできるようになったことが極めて大きい。しかもネットワークの世界はドッグイヤーと呼ばれるようにその急成長ぶり、つまりこの場合ホームページ(HP)サイトの充実ぶりは表現を超えるものがある。現在では貴重な言語学資料のほとんどが入手できるほどである。即ち、インターネットは過去から現在に至るまでの英語情報の宝庫である。(もちろん、過多ともいえる情報のなかから必要な情報を検索・収集し、分析する能力が問われることにもなる。)こうしたインターネットの普及、HPの充実等々は、英語コーパス言語学の必要性を増大させ、その可能性を限りなく広げている。そこで次に、英語コーパス言語学の内実を考察する。

二 コーパスとは何か？

まず、コーパス (corpus) とは何だろうか？ 現在ではコーパスとはコンピュータ・コーパス若しくは電子コーパスを意味するが、コンピュータ以前にもコーパスは存在した。コーパスとは「もともとラテン語で「体」を意味する語であったが、すでに古代ローマ時代から文献や事実の集大成を意味する語として使われていた。この語はそのままの形で学術用語としてヨーロッパの多くの言語に取り入れられ、この用法は現代まで引き継がれている。」⁽³⁾ 従って、コーパスとは、コンピュータ利用以前の一般的理解としては、ある特定のテキスト群の資料総体といえるだろう。

さて、コーパスは言語学上の視点とデータ保存及びテキスト収集の視点とにより、広義と狭義のコーパスに区別される。*The Oxford English Dictionary* (second edition, 1989)にはコーパスを次のように定義している。“The body of written or spoken material upon which a linguistic analysis is based.” *OED*に拠ればこの定義の初例は、1956年である。広義の定義としては、この「言語分析が依拠する書かれたり話されたりする言語資料の集積」という定義が共通理解であると思われる。次に、コンピュータの保存機能の視点から見れば、「言語研究に使用されることを想定して、実際に書かれたり話されたりした言語をコンピュータ上で利用可能にしたテキストの集合体」⁽⁴⁾という狭義の定義が成り立つ。また、テキスト収集の視点から見れば、「言語研究のために研究対象となる言語変種(language variety)を代表するように、明確なコーパス・デザインをもって収集されたテキストの集合体」⁽⁵⁾という狭義の定義が成り立つ。従って、両者を組み合わせて、狭義のコーパス定義とは、「ある特定の言語、方言、もしくはその他のヴァリエティを代表し、かつ言語研究に使用されることが想定され、コンピュータに蓄積、処理された話し言葉、書き言葉のテキストの集合体」⁽⁶⁾ということになるだろう。

これに対し、後藤齊氏は前出の論文において、ヨーロッパ英語圏の詳細なコーパス定義検証を行っているが、G.Leechのいろいろ変わるコーパス定義の真意を推測し、「広義では、... コーパスとはコンピュータで扱われる大規模なテキストの集積であって、... 狭義では、大規模なテキストの集積(すなわち、広義のコーパス)のうちで、言語学の研究を目的としてある一定の方針の下に集められたものだけがコーパスと呼ばれる」⁽⁷⁾、とかなり相違のある見解を紹介しておられる。こうしたコーパス定義理解に相違が生じてしまうのは、コーパス言語学そのものが比較的新しい学問であり、かつコーパスがもつ特性とそこから生じるコーパスの種類に原因があると思われる。とまれ、コーパス言語学(corpus linguistics)とは、コンピュータで処理可能な(machine-readable)言語資料(corpus)を用いた新しい言語学であり、英語コーパス(English corpus)を用いた英語研究法が英語コーパス言語学である、と定義づけしても一般的理解の合意は可能であると思われる。

それではコーパスの特性とは何だろうか？ それは一言でいえば、代表性(representativeness)にある。例えば、アメリカ英語とイギリス英語の話し言葉を比較検討する場合に、そのすべてをコーパスに収めることはできない。そこで、両者のサンプルを抽出することが重要になる。その際、それらのサンプルが夫々の英語を代表していなければ相互の比較検討は成立しない。つまり、コーパスには、収集されたテキストが研究対象となる言語変種を代表している、という代表性がなければならぬ。梅咲敦子氏は、この特性に従ってコーパスを分類されておられるが、要約すると以下の通りである。

1 サンプル・コーパスとモニター・コーパス

サンプル・コーパス(sample corpus)とは、一定量のサンプルを電子化したコーパスのことであり、一般的にコーパスといえは、サンプル・コーパスのことである。有名なBrown CorpusやLob Corpusはその典型である。それに対して、常に変化する言語を監視しながら、古い情報を捨て、新しい情報を付け加えて行くコーパスがモニター・コーパス(monitor corpus)である。Bank of Englishを構成する一部分はこのコーパスを代表している。

2 特殊目的コーパスと汎用コーパス

サンプル・コーパスは大きく二つに分類される。一つは、特定の言語研究のために編纂される特殊目的コーパス(special purpose corpus)であり、もう一つは、総合目的のために編纂される汎用コーパス(general purpose corpus)である。前者の代表例として、子供の話し言葉を集めたPoWがある。また、Brown CorpusやLob Corpusは、後者に属する。

3 共時コーパスと通時コーパス

サンプルを同時代に限って抽出したコーパスが共時コーパス(synchronic corpus)であり、複数の時代区分を設け、その時代別にサンプルを抽出して整理したコーパスが通時コーパス(diachronic

corpus)である。前者にはBrown CorpusやLob Corpusが含まれ、後者には1650～1990年間のテキストを集めたARCERがある。この他にも言語媒体の視点から分類すると文字言語コーパスと音声言語コーパスとに分類できる。前者は、書き言葉からサンプルを抽出したコーパスであり、後者は、話し言葉を録音し文字化したコーパスである。⁽⁸⁾

更に、例えば、バイリンガル地域であるカナダのケベック州の議会文書は英語とフランス語で書かれているが、このようにまったく同じ内容の2カ国語コーパスをパラレル・コーパスという。また、そっくり同じに翻訳するのではなく、似かよったテキストからできている二つの異なる言語のコーパスをコンパラブル・コーパスという。世界15カ国で使われている英語のコーパスから構成されるICE（正式名はInternational Corpus of English）はその代表例である。⁽⁹⁾

このように英語コーパスの特性及びそれに基づく分類をみてくると、伝統的英語学観のカテゴリーと重なり合う箇所が多いことに気づく。何故か？なぜなら、それは両者が英語という特定言語を対象にした科学的研究という始源を同一にしている言語学であるからである。それ故、「英語コーパス言語学は英語学の一分野なす」⁽¹⁰⁾と既に認知されていると思ってよいだろうし、また、今後英語学の中心分野の一つになることが期待される。しかし、それが、音韻論や形態論、統語論といった従来の英語学観の中心をなす言語固有の研究分野をもつということではなく、コーパスに基づく音韻論や形態論、統語論をもつということである。つまり、英語コーパス言語学は従来の領域にコーパスを活用して行くツール若しくはパイロットになる方法論である。更に、或いは言語学の各分野の視点からコーパスを活用して英語を科学的に研究する学問といえるだろう。齊藤俊雄氏は「それ（＝コーパス言語学）は電子コーパスによる言語研究の方法論(methodology)であり、コンピュータ・テクノロジーのハードウェアとソフトウェアの発展に支えられて、……言語学の各領域および関連分野に活用されて、豊かな成果を生み出す可能性を持ったものである」⁽¹¹⁾と指摘しておられる。

三 英語コーパス言語学の歴史的背景

(一) Brown Corpus

コンピュータ以前にコーパスを利用した最初の言語学的研究は、1870～1926年にかけて行われた子供の言語獲得状況に関する親の育児記録といわれている。次に1897年にJ.W.Kaedingがドイツ語の文字頻度分布とその連続性に関する1100万語のドイツ語コーパスを作成した。その後1940年にFriesとTraverが、1947年にBongresが手作りのコーパスを外国語教育における語彙研究に活用した。これらは膨大な時間と手間をかけての手作業と肉眼によってカウントされながら作られたコーパスである。⁽¹³⁾ところで、明確なコーパス・デザインをもって編纂された最も古いコーパスは1959年に着手され1960年代に発表された R.Quirk とS.GreenbaumとによるSurvey of English Usage Corpus (SEUコーパス)といわれている。このコーパスはイギリスの話し言葉（独白、対話、インタビュー等々）

100万語、書き言葉（印刷物、ニュース原稿、演説原稿等々）100万語からなる計200万語のコーパスである。これも一切手作業によって編纂された。このコーパスの話し言葉の部分は1975年に電子化されて現在でも話し言葉コーパスのモデルとなっているLondon-Lund Corpus of Spoken Englishとなった。また、SECコーパスから数多くの実証的研究が生じ、現代最高の英文法書といわれているA Comprehensive Grammar of the English Language (1985)はこのコーパスを母体としている。⁽¹⁴⁾

さて、同時期（1961～1964年）に米国のブラウン大学のW.N.FrancisとH.Kuceraとにより編纂され、現在でも幅広く利用されているコンピュータ・コーパスが、Brown Corpus（正式名は、The Brown University Standard Corpus of Present-Day Edited American English）である。このコーパスは、今日に至るまで英語コーパス言語学の在り方を規定し続けているサンプル・コーパスである。そのコーパス・デザインは、母集団として1961年にアメリカで印刷刊行された散文の文字言語を新聞雑誌の報道・論説・評論、文学、宗教等々の15のジャンルに分類する。その夫々を代表する約2000語のテキスト500を集めて100万語にしたコーパスである。2000語目を含む文の終わりまでを採用していたため内容上必ずしも完結してはいない。サンプル自体は作業上の問題もあって必ずしもアットランダムに選定されたわけではない。その特徴は、「一定の観点から分けたジャンルにそれぞれ重みをつけ、その重みに比例した数の比較的小さなサンプルを無作為に選び、それによって母集団を代表させた」⁽¹⁵⁾ ところにある。こうしたBrown Corpusのデザインはその後のコーパスのモデルとなり、例えば、ランカスター大学のG.Leechによって1970年に着手され、オスロ大学のS.Johanssonとベルゲン大学のK.Hoflandによって1978年に完成したイギリス英語初のLOB Corpus（正式名はThe Lancaster-Oslo / Bergen Corpus）がある。まさにBrown Corpusによって初めて、「アメリカ英語の本格的な統計的研究が可能になり、コンピュータ利用のコーパス言語学の誕生を見た」⁽¹⁶⁾ といえる。Brown CorpusやLOB Corpusはペーパーバックの本にすると約数十冊分の本に相当する分量である。⁽¹⁷⁾ 1980年代以降、従来大型コンピュータで処理していたコーパスが、パーソナル・コンピュータの開発とマシンの能力(CPU,ハードディスク、メモリー等々)の向上により、億語単位のコーパスができるようになった。このことは100万語規模のコーパスは100万語規模で十分な研究には機能を発揮したが、それ以上の語彙数とサンプル数を必要とする研究には不十分であった。例えば、Brown Corpusの異なり語数は約5万語であり、このうち多くの語は一度しか現れない。このことは語彙研究としてはデータ不足である。大規模コーパスとしては、1991年に着手され1994年に完成したThe British National Corpusがある。これは、1975年以降のイギリス英語を対象にし、書き言葉9000万語、話し言葉1000万語、計1億語のコーパスである。また、1991年から始まり1995年に終了したBOE(Bank of English)は、現在世界最大のコーパスといわれている。このコーパスは、英米の書籍、雑誌、話し言葉、BBC、NPR(National Public Radio)、イギリスとオーストラリアの新聞、カタログ、などからなる約3億2900万語のコーパスである。このうち5000万語はCobuilddirectサービスによってインターネットから利用するものである。⁽¹⁸⁾ また、ハーパーコリンズ社の辞書はこのBOEに依拠している。綿密なコーパス・デザインに基づいて作成されるサンプル・コーパスに対して、コンピュ

ータ・テクノロジーの展開と相まって、多様化する社会の変化のなかで言語の全体像を大まかに捉えるために語彙・テキストの拡張を図るコーパス、即ち、モニター・コーパスの構想が生まれてくる。それは量の相違が質の相違を生じるという考え方である。Collins COBUILD Dictionaryの編集主幹であったSinclair等によって提唱されるようになった。この辞書の730万語のMain Corpusはサンプル・コーパスであるが、それを補足する1300万語のReserve Corpusはモニター・コーパスの構想を表示している。

さて、上記の代表的な幾つかのコーパスの考察によってもある程度知れることであるが、コーパス言語学の特徴は、以下の4点に集約できるかと思う。

1. 言語能力(competence)よりも言語運用能力(performance)に中心を置く。
2. 言語の普遍的特性(linguistic universals)の解明よりも個別言語の言語記述(linguistic description)に中心を置く。
3. 質的な(qualitative)言語モデルのみならず、数量的な(quantitative)言語モデルにも中心を置く。
4. 言語研究における合理主義的(rationalistic)な立場よりもむしろ、経験主義的(empirical)な立場に中心を置く。⁽¹⁹⁾

これらの特徴により、一見して英語コーパス言語学は、1からは実践的であること、2からは現実的であること、3からは帰納的であること、4からは経験主義的であることが理解される。

(二) 経験主義と合理主義

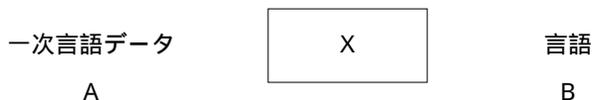
英語コーパス言語学は、言語研究史において合理主義(rationalism)の立場ではなく、経験主義(empiricism)若しくは実証主義(positivism)の立場に依拠する。経験主義とは、経験を超える知識を知識と認めない。つまり、ア・プリオリな認識を拒絶する思想である。権威化した形而上学への批判といえる。それは、ロック、ヒューム、ベンタム、そしてJ.S.ミル等々のイギリス経験論哲学者たちによって確立された。また、このイギリス経験主義やフランス啓蒙主義の流れに帰属する実証主義は、確かな知識獲得の在り方として自然科学的方法である実験や観察に依拠する立場をとる。実証主義者たちは、この立場を文化・社会科学の領域に導入し、従来支配的であった神学や形而上学的見方を転換しようとした。サン シモン、コント、デュルケム等々がこの思想を代表し、コントは人間の知識の発展を神学的、形而上学的、そして実証主義的段階に3区分し、認識は経験の事実面に即していなければならない、と主張した。⁽²⁰⁾ また20世紀に入るとF.deソシュール(1857-1913)にはじまり、プラハ学派のR.J.ヤコブソン、C.レヴィ ストロース(1908-1992)等々を経て展開してきた構造主義(structuralism)が言語学界で注目を集めるようになった。ソシュールは言語学の当時の主流であった歴史的研究から構造的な研究に転換し、パロ - ル(話される言葉)の恣意性とラング(記号システムとしての言語)の共時性を指摘し、両者を区別した。言語と指示対象との間には何ら必然性

はなく、従って、言語はコミュニケーションの手段ではなく、世界を分節して理解するためのメディア若しくはツールと考えたのである。また、アマゾンの未開人社会の現地調査を行い、人間社会の基底的構造モデルを探究したC.レヴィ ストロースは、1960年代に当時フランスで主流であった現象学や実存主義の主体性を否定し、人間経験の中心を主観的意識から構造へと脱中心化していった。これらイギリス、ヨーロッパ大陸の思想はアメリカ先住民の言語や文化を研究していたアメリカの言語学学界にも影響を与えた。20世紀半ばから始まるアメリカの構造言語学(structural linguistics)は経験主義の言語学であり、主としてL. BloomfieldやC.C.Fries等々によって発展した。アメリカ構造言語学を端的に説明すれば、それはコーパスを分析や記述の対象とし、そこに観察される言語の特徴や規則性を記述する言語学である。この立場は機械的、帰納的、そして分類的な行動主義心理学(behaviorist psychology)を発展させることとなるし、逆に行動主義心理学がアメリカ構造言語学の隆盛を招くことにもなった。⁽²¹⁾

行動主義心理学とは、人間の行動を内部からではなく、外部から観察される現象として分析する行動主義(behaviorism)の心理学である。J.B.WatsonやB.F.Skinnerが代表的であるが、彼らは従来の意識心理学の立場を転換し、行動はア・プリオリではなく、習得によって生じると考えた。つまり、行動は刺激とそれに対する反応に依拠すると考えた。例えば、刺激(「煙草を買ってきて」という言葉)に対する反応(「煙草を買いに行く」という行動)がその刺激の意味であると考えている。彼らは複雑な内面的心理よりも、客観的に観察できる言語活動の分析を任務としていた。⁽²²⁾「意識よりも行動を、内観よりも客観的観察と実験を、本能と遺伝よりも学習と環境を重視した」⁽²³⁾といえる。こうしたアメリカ構造言語学や行動主義言語学を背景に大型コンピュータを駆使して作られたコーパスが、以後のコーパスを規定することになる既出のBrown Corpusである。しかしこれ以後アメリカではコーパス言語学は衰退して行く。何故か？ 合理主義に基づくチョムスキーの生成文法(generative grammar)が台頭し、コーパス依存の言語研究は疎外されたからである。

チョムスキーは第一義的に「言語能力」と「言語運用」とを区別し、言語学が取り組む問題は前者であるとした。既にみたとおり、コーパス言語学は後者を主として扱う。チョムスキーの「Syntactic Structure(1957)に始まる生成文法では、言語分析は母語話者の言語的直観に依存する言語能力の解明で」⁽²⁴⁾あった。従って、チョムスキーの生成文法の着想は経験主義に依拠するアメリカ構造言語学や行動主義言語学の批判から生まれたともいえるかもしれない。経験主義はしばしば人間はタブラ・ラサで生まれてくると考え、生得的なものを否定するが、彼は生得的な文法能力(普遍文法)を認めるのである。言語の本質的な部分はア・プリオリに与えられており、社会や環境から独立していると考えるのである。例えば、人間の母国語習得の場合を考えてみる。その場合、3つの項が考えられる。第一の項は新生児が成人するまでの間に周囲の人々から受ける一次言語データ(これをAとする)であり、第二の項は母国語を自由に操れる能力を身につけた状態(これをBとする)であり、第三の項がAを処理してBを生ずる媒介項というべきBlack Box(これをXとする)である。これは未知項である。AとBは既知項である。Aを入力して、Xのフィルターを通して、

Bを出力するモデルは下図のように表示される。



ではこの図式が成立するためには、Black Box Xにどのような能力があればよいだろうか？Xが空っぽであれば母語の習得は成立しない。従って、チョムスキーはBlack Box Xに母語習得の生得的な能力を認めるのである。しかも、Aの無限の構成要素に対してXの処理を通してBの無限の出力結果が得られる。つまり、我々は必要に応じていくらでも新しい文を創造でき、またそれを理解できる能力をもつ。彼は人間言語の創造性に言語研究を主題化していったのである。仮説ではあるが、この彼の生成文法がどの言語にも妥当するとすれば、即ち、Black Box Xのメカニズムにどの言語にも妥当する性質があるとすれば、それは普遍性をもつことになる。従って、彼は一つの言語、つまり、英語だけを研究しても普遍文法に到達できると考えた。彼は生得的言語能力が深層構造・普遍文法の基底を形成すると考えていたのである。⁽²⁵⁾ こうしてアメリカで衰退したコーパス言語学は、その中心地が経験主義・実証主義の伝統が強いイギリス、ヨーロッパ大陸へと移った。そして、Brown Corpus成立から14年の歳月の経過をみた後、LOB Corpusが完成した。

四 結 語

チョムスキーに批判され、一時期下火になった英語コーパス言語学であるが、両者は言語学を科学とみる点では一致している。相違するのは、先ず、その方法論である。チョムスキー自身データ重視の構造言語学から出発したが、彼はその帰納的方法論に限界を感じ演繹的方法論へ赴いた。彼は、ある文の文法的正否は発話の使用頻度の統計的確率ではなく、発話を解明する理論こそが必要と考えた。例えば、“I live in New York” という文が “I live in Dayton, Ohio” という文より、その頻度が高いのは当たり前ではないか、とある講演で述べたコーパス言語学批判は有名なエピソードである。彼にとって生得的な普遍文法が問題であり、彼はそこから派生する現象の分析では本質に至らないと考えたのである。この相違から必然的に科学の理解に対する相違も生まれる。チョムスキーは、我々の脳に内在する文法規則を探究するために母語話者の直観を基準とする。従って、言語の実際的運用情報、即ち、コーパスによるデータに意義を認めないのである。⁽²⁶⁾ しかし、我々人間が間主観的な存在として言語をもって社会にその生を営んでいる以上、一方的に社会と言語を切断することが可能であるだろうか？ 経験主義と合理主義（普遍主義）とは、水と油のように永遠の相克の関係にあるように思われるが、両者の止揚が今後の言語学、英語教育への貢献の立場からすれば最も望ましいのではないかと思われる。今後、この点についての研究も進めてみたい。

（よしの きよし・本学非常勤講師）

註)

- (1) 鷹家秀史・須賀廣『実践コーパス言語学』、桐原ユニ、1998年、pp.39 - 40参照。
- (2) 増田貢『現代英語学』、篠崎書林、1972年、pp.1 - 14参照。
- (3) 後藤斉「言語研究のためのデータとしてのコーパスの概念について 日本語のコーパス言語学のために」、『東北大学言語学論集』第4号、1995年、<http://www.sal.tohoku.ac.jp/gothit/corpus.html>。
- (4) 梅咲敦子「コーパスとは何か」、齊藤俊雄、中村純作、赤野一郎編『英語コーパス言語学 基礎と実践』、研究社、1998年、p.17。
- (5) 同書、p.17。梅咲敦子氏は、言語変種に「言語を使用されている社会階層、地域または使用域により分類した区分」との注を付けておられる。
- (6) 同書、p.18。
- (7) 後藤斉「言語研究のためのデータとしてのコーパスの概念について 日本語のコーパス言語学のために」、<http://www.sal.tohoku.ac.jp/gothit/corpus.html>。
- (8) 梅咲敦子「コーパスとは何か」、p.18 - 20参照。
- (9) 鷹家秀史・須賀廣『実践コーパス言語学』、桐原ユニ、1998年、p.72参照。
- (10) 齊藤俊雄「英語コーパス言語学とは何か」、『英語コーパス言語学 基礎と実践』、p.3。
- (11) 同書、p.4。
- (13) 鷹家秀史・須賀廣『実践コーパス言語学』、pp.29 - 30参照。
- (14) 齊藤俊雄「英語コーパス言語学とは何か」、pp.5 - 6参照。
- (15) 後藤斉「言語研究のためのデータとしてのコーパスの概念について 日本語のコーパス言語学のために」、<http://www.sal.tohoku.ac.jp/gothit/corpus.html>。
- (16) 齊藤俊雄「英語コーパス言語学とは何か」、p.6。
- (17) 何れのコーパスともICME(International Computer Archive of Modern and Medieval English)からダウンロードできる。URLは<http://hd.ubi.no/icame.html>である。
- (18) 滝沢直宏『電子化コーパスの利用』
<http://lang.nagoya-u.ac.jp/takizawa/kokai/corpus/text1.html>参照。
- (19) 齊藤俊雄「英語コーパス言語学とは何か」、p.4。齊藤俊雄氏によるG.Leech著、“Corpora and Theories of Linguistic Performance”(1992), in Svartvik (ed.), pp.105 - 22の要約である。
- (20) 片山洋之助「科学技術と人間」、星野勉『『合理性』をめぐって』、城塚登・片山洋之助・星野勉『現代哲学への招待』、有斐閣、1995年、pp.84 - 85, pp.124 - 125参照。
- (21) 田中春美編集『現代言語学辞典』、成美堂、1988年、p.61, pp.637 - 640参照。
生松敬三・木田元・伊藤俊太郎・岩田靖夫編『西洋哲学史の基礎知識』、有斐閣、1984年、pp.295 - 296参照。
井関利明『60年代におけるパラダイム変革運動』、1994年、
<http://www.dnp.co.jp/mediascape/no01/s04.html>参照。
- (22) 片山洋之助『『私』をめぐって』『現代哲学への招待』、pp.26 - 27参照。
- (23) 濱田庸子『人間行動論』、<http://www.sfc.keio.ac.jp/%7Et94208ns/human/1.html>。
- (24) 齊藤俊雄「英語コーパス言語学とは何か」、pp.6 - 7。
- (25) 安井稔『英語学の世界』、大修館書店、1974年、pp.215 - 228参照。
福井直樹「言語の普遍性と多様性」、『言語』第28巻第12号、1999年、pp.22 - 32参照。
- (26) 鷹家秀史・須賀廣『実践コーパス言語学』、p.30, pp.33 - 36参照。
後藤斉「言語研究のためのデータとしてのコーパスの概念について 日本語のコーパス言語学のために」、<http://www.sal.tohoku.ac.jp/gothit/corpus.html>参照。