様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ５ 月 ２６ 日現在

機関番号：１４３０１
研究種目：基盤研究(C)
研究期間：2010～2013
課題番号：２２５２０４９５

研究課題名（和文）英語の史的コーパス構築とその利用による歴史社会言語学的研究

研究課題名（英文）Corpus-based Methods in English Historical Sociolinguistics

研究代表者
家入　葉子（IYEIRI, Yoko）

京都大学・文学研究科・教授

研究者番号：２０２６４８３０

交付決定額（研究期間全体）：（直接経費）　3,000,000円、（間接経費）　900,000円

研究成果の概要（和文）：　4年間の研究期間に、パイロット版のEarly Modern English Prose Selections (EMEPS)（1500年から1700年の英語の史的コーパス、約800万語）を構築し、その有用性を確認するために、これを利用して、英語の動詞の構文の発達を中心に言語分析を行った。研究から、さまざまな動詞が不定詞や動名詞の構文を徐々に発達させていく過程を明らかにすることができたと同時に、EMEPSをほかのコーパスと組み合わせて利用することにより、より有効な分析が可能になることを明らかにすることができた。

研究成果の概要（英文）：I have compiled a collection of electronic texts (version 1, around eight million words), which I tentatively call Early Modern English Prose Selections (EMEPS), and conducted research in to it during the past four years. This is to prove that building corpora for one's research purposes is a useful method in historical sociolinguistics and linguistics in general. EMEPS is not a balanced-corpus, but in combination with various single-genre corpora, it can effectively be used to highlight different stylistic and sociolinguistic aspects in the development of the English language. In the present project, I focused upon patterns of complementation of various English verbs, e.g. doubt, make, and prohibit, and discussed how they changed during the Early Modern English period (c.1500-1700).

研究分野：人文学

科研費の分科・細目：言語学・英語学

キーワード：歴史統語論　データベース　初期近代英語　歴史社会言語学　英語史　中英語　コーパス

１．研究開始当初の背景

20 世紀の後半以降、英語の語学的研究、特に実証的な研究においては、電子化されたテキスト、すなわちコーパス（＊）を利用することが一般的な方法論として確立しつつある。1991 年にヘルシンキコーパスが公開されてからは、英語の史的研究の分野でも、コーパス利用による研究が急速に発展してきた。

本研究プロジェクトが開始された 2010 年は、ヘルシンキコーパスの公開からほぼ 20 年が経過したところであり、コーパス言語学の方法論そのものの展開を概観するとともに、新たなコーパス言語学の研究方法を模索する時期に入っていたということができる。

（＊電子化されたテキストの使用が言語研究の本格的な方法論として確立してくる以前には、「コーパス」という用語は、言語研究のための資料の意味で使用されていた。この段階でのコーパスは必ずしも電子化を前提としていなかったが、現在では、電子化されていることは、「コーパス」の中心的な定義の一つとなっている。以下では、「コーパス」という用語を現在の意味で使用することとする。）

２．研究の目的

上述のように、コーパス言語学は 20 世紀の後半以降、急速な拡大を経験してきた。この間の研究動向としては、コーパス言語学の方法に基づいた研究そのものが増加しただけではなく、コーパス言語学のあり方そのものも変化しながら発展してきたといえる。特にこの傾向は現代英語のコーパスの性質の変化に顕著であり、初期の 100 万語レベルのコーパスから、コーパスが億の単位の語数を含むような大規模化を経験し、一方で、文法的な検索などを可能とするタグ付けの技術が進んできたことなどをあげることができる。

このような中で、多くの研究者が一般公開のコーパスを共有しながら使用する傾向が強まり、特に簡便なウェブ・コーパスの使用の拡大には著しいものがある。

研究代表者は、このようなコーパス言語学の動向を肯定しながらも、一方で、多くの研究者が同一のコーパスに集中する状況に疑問を感じ、コーパス言語学の今後の展開の可能性を探ることを本研究の主要な目的とした。現在は、コーパスに限らず、電子テキストがさまざまな形で入手可能になるとともに、研究者自身が文献を電子化してコーパスを作成することも、比較的容易になってきている。これまでのような大規模の研究プロジェクトに依存するのではなく、個人研究のレベルでもコーパスの構築が可能になってきていることを踏まえ、実際にコーパスの作成を行うことで、コーパス言語学のさらなる可能性を追求しようとする試みである。コーパス言語学が、新たな方向への発展を示そうとしているのではないか、という研究代表者の直観からスタートしたプロジェクトである。

３．研究の方法

（1）研究の方法は複雑なものではなく、個人レベルでのコーパスの作成とそれを利用した言語研究とから構成されている。ただし、その中間の段階として、いったん作成したコーパスを利用しながら、コーパスの有用性を確認する、という作業を含んでいる。

（2）上記の（1）を行うために、特に研究の対象とする時代としては初期近代英語期（1500 年頃から 1700 年頃）を選んだ。英語史は 5 世紀の半ばにゲルマン人がブリテン島に移住したところに始まり、現在まで 1500 年以上にわたって継続しているが、その中で、特に変動が激しかった時期の一つとして初期近代英語期をあげることができる。本研究では作成したコーパスの検証という作業も重要な位置を占めており、言語変化が大きな時期は、この作業に適しているということができる。時代ごとの変化がコーパスの分析結果からきれいに示されるか、が判断の手掛かりになるからである。ただし、これらの作業はコーパス言語学全般の枠組みの中で進めることを前提としているため、必要に応じて、初期近代英語以外の時代にも目を向けることとした。特に中英語（1100 年～1500 年）後期は、初期近代英語期に並んで言語変化が大きな時期であると同時に、初期近代英語の流れを作る重要な時期でもある。また、現代英語はコーパス言語学の発展の中心的な役割を担っている。そこで、これらの時代についても視野を広げながら研究を進めることとした。

（3）本研究の顕著な特徴の一つは、すでに電子化されて入手可能となっているテキストを利用しながら、研究者自身の研究目的のために比較的大きなコーパスを作成した点である。利用したのは有料で公開されているデータベース、Early English Books Online である。中英語後期から初期近代英語期にかけても電子テキストを含んでいるが、この中から 1500 年～1700 年の資料を検討し、約 800 万語の Early Modern English Prose Selections (EMEPS)を作成した（ただし、現段階ではパイロット版）。有料のデータベールをもとに集積した電子資料であるので、著作権の関係で、あくまで研究者個人の利用にとどめなければならないという制約はあるが、コーパスそのものではなく、コーパスの情報や作成方法については、今後も公開していく予定である。

（4）最後に（3）で作成した EMEPS を利用し、さまざまな言語研究を行った。必要に応じてその他の電子テキストや一般公開の汎

用性の高いコーパスの利用も行い、その研究結果との比較等も行った。

４．研究成果

本研究の成果は、電子テキストの構築、それを利用した言語分析、そして言語分析を通してコーパスの構築のあり方を検討すること、の 3 つの分野に分類することができる。ただし、これらは互いに関係しており、厳密な分類は容易ではないので、あくまで大まかな分類である点を断っておきたい。以下、それぞれについて詳細を記す。

（1）コーパスの構築については、さまざまな電子テキストの集積とその利用を試みたが、その中でもっとも大きな成果といえるのは、上述のように EMEPS のパイロット版の一応の構築を終えたことである。約 800 万語からなる EMEPS は、A テキスト 400 万語、B テキスト 400 万語の 2 種から構成されており、それぞれ 1501～1550 年、1551～1600 年、1601～1650 年、1651～1700 年の 4 期に分かれている。頻度が高い言語現象を分析する場合には、A テキストのみ、あるいは B テキストのみを分析の対象とし、頻度が低い言語現象を調査する場合には、A テキストと B テキストの両方を扱うことができるよう、構成に工夫を加えている。このため、研究目的に応じて複数の利用方法が可能となっている。

（2）次に、（1）のコーパスや電子テキストを利用しながら言語分析を行う点についてである。言語分析では、初期近代英語期を中心に、その前後の時代にも気配りをしながら、研究代表者が近年関心をもっている動詞およびその周辺領域について、いくつかの研究成果を発表することができた。たとえば、prohibit、make などの動詞の構文は歴史的に大きな揺れを経験しながら発達してきた経緯があり、これらの動詞について、その発達の過程を明らかにすべく、量的な分析を進めた。prohibit は現在では「prohibit＋人＋from *–ing*」のような構文で使用されるが、歴史的には that 節や不定詞を従える構文が可能であったことが知られており、この変化を追跡するためには、数百年間にわたる言語データの解析が必要である。本研究では、初期近代英語期を中心に、その一端を明らかにすることができた。

一方、make については、その使役動詞としての用法の発達が興味深い。現在では、能動態では原形不定詞を伴い受動態では to 不定詞を伴うとされているが、この使い分けも、数百年という時間をかけながら、主に近代英語期に確立してきたものである。本研究ではまず、この使い分けが徐々に確立してくる前提条件を明らかにするために、中英語後期の make の分析をおこなった。なお、上述の EMEPS では対応できなかったので、インスブルックコーパスなど、利用可能な電子テキストを組み合わせながら研究を進めた。その結果、中英語後期の make が原形不定詞を選択するか to 不定詞を選択するかには認知等のさまざまな要因が関係していることがわかり、その後の発達についての推論を行う上で重要と考えられる情報を得ることができた。

以上のほかにも、doubt や convince など、複数の動詞について、同様の分析を進めるとともに、動詞に関係すると考えられる言語事象についても、一部研究を開始したところである。副詞の always の発達はその一つで、中英語期までは圧倒的に alway のような-s がつかない形がふつうであるが、これが初期近代英語以降、急速に語尾の-s を獲得し、形態として always を確立させていく。EMEPS の分析を通して、この変化が特に顕著であるのは 16 世紀であることが明らかになった。16 世紀の前半ではまだ、両者の割合がかなり拮抗しているのに対し、世紀の後半になると、明らかに always の頻度が優勢になる。そして 17 世紀になると、always がほぼ確立する勢いである。always の形態に関する研究は、これまでほとんど先行研究がなく、全体的な発達の方向は英語史研究者の間で知られていたものの、その過程を明らかにする記述はほぼ皆無であった。本研究で、少なくともその発達過程の一部を明らかにすることができたといえよう。

このほか、本研究では副詞的な節を導く接続詞と語順の問題についても研究を進めたが、この分野については、近代英語の分析の準備段階として、中英語後期、特に『パストン家書簡集』の分析を行ったにとどまっている。語順も動詞とかかわる重要な分野であり、今後、この成果を近代英語期以降の分析にも生かしていきたいと考えている。現段階で明らかになった点は、同じ副詞節でも、どの接続詞が用いられるかで、主節と従属節の配置に大きな違いが見られ、たとえば条件を表す if 節の場合は主節の前に来ることが多く、時間を表す till のような従属節は、時間関係を反映しながら主節の後に来ることが多い、などの特徴がみられる。人間の認知のメカニズムが言語のあり方に反映する傾向が強いことが明らかになった。この点は、先に述べた個別の動詞の構文等にも応用できる可能性が高く、これら一連の研究を進めていくことで、将来的には、言語の発達と認知の仕組みの関係を総合的に明らかにできるのではないかと考えている。

（3）最後に電子テキストの集積によるコーパスの構築等にかかわる有効性の確認と方法論の部分についての成果に言及したい。より具体的には、上述の（2）の言語分析の結果をどのような形で（1）の評価につなげたかという点である。本研究ではいくつかの電子テキスト（あるいはグループ化された電子

テキスト）を利用したが、その有効性の検討については、主に EMEPS を対象とした。いくつかの言語分析から明らかになった点は、A テキスト、B テキストのように自由な組み合わせを許す構造をもち、全体としては、少なくとも英語史分野のコーパスとしては比較的大きいといえる EMEPS ではあるが、動詞 prohibit のような頻度が低い動詞では、やはり規模の点で、やや不足を感じざるをえなかったことである。一方、副詞 always のような高頻度語については明らかな傾向を指摘するのに十分な言語データを、比較的容易に得ることができた。コーパスの規模をどのようにデザインするかについての研究は多いが、この問題が、扱う言語現象、すなわち研究目的に大きく依存することが、本研究でも再確認できたといえる。大規模コーパスをウェブ上に公開する Mark Davies（http://corpus.byu.edu/参照）は、規模の大きさがコーパスの威力につながる点を確信し、その規模を継続的に拡大させている。一方で、Marianne Hundt と Geoffrey Leech のように、規模を押さえることのメリットの言及する研究者も少なくない（‘Small is beautiful’: On the Value of Standard Reference Corpora for Observing Recent Grammatical Change, in *The Oxford Handbook of the History of English*, ed. Terttu Nevalainen & Elizabeth Closs Traugott (OUP, 2012), pp. 175-88 参照）。この領域については、今後も検討が必要であろうが、最終的には研究の目的次第というところがあり、ここでも個人の研究者が必要に応じてコーパスを作成することの意義が確認されたということができる。

最後に本研究では、コーパスの構築とジャンルの問題について検討した。EMEPS はジャンルの構造を意識して作成したコーパスでは必ずしもない。一定の量を集積することにより、全体としての傾向を示し、これと特定のジャンルに特化した特殊コーパスを組み合わせることでジャンルの問題にも議論を発展させることができるという想定のもとに構築された電子テキストの集積である。この意味で EMEPS は、いわば参照コーパスであるということができよう。上述の（2）の always の研究では、特にこの点の検証を意識的に行った。より具体的には、EMEPS の研究結果と Merja Kytö と Terry Walker の作成による特殊コーパス A Corpus of English Dialogues 1560-1760 の分析結果を比較するという試みを行ったが、その結果両者の違いがきわめて明確な形で明らかになり、always の発達が話し言葉から先に起こり、書き言葉の変化を促した可能性を指摘することができた。また、この結果をヘルシンキコーパスの研究結果とさらに比較することで、EMEPS のこのような利用方法そのものの有効性を、ある程度確認することができたといえる。ヘルシンキコーパスにおいても、法律などの文書では always のように-s がついた形態の発達が遅く、一方で同じ法律のジャンルでも、話し言葉を反映している（あるいは話し言葉により近い）と考えられる裁判記録などでは always の発達が早い。コーパスをどのように組み合わせていけば、より有効な利用が可能になるかについては、今後さらなる検討が必要であると考えられるが、少なくとも always の分析から、この手法の有効性がある程度見えてきたということができよう。

以上のように、本研究では、コーパスの構築、それを利用した言語分析、そして言語分析を通じて、一定の成果を収めることができたと考えられる。一方で、今後の研究のための課題と方向性も明らかになってきた。EMEPS は一つの成果ではあるが、研究の目的によっては対応できない点もあることが明らかになったので、今後も別な形での電子データの集積も行う必要があろう。言語分析については、動詞を中心に進めてきたが、まだ分析の対象とした動詞が限られており、また動詞によっては、一部の時代の分析で終わったものも少なくない。さらなる動詞の分析を進め、また時代を広げていくことで、結果的にコーパス言語学の方法についての検討もさらに進むことになるであろう。いずれも今後の課題である。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　2 件）

① 家入　葉子、Early Modern English Prose Selections: Directions in Historical Corpus Linguistics、京都大学文学部研究紀要（Memoirs of the Faculty of Letters, Kyoto University)、査読無、50、2011 年、133－199 http://repository.kulib.kyoto-u.ac.jp/dspace/bitstream/2433/139204/1/lit50_133.pdf

② 家入　葉子、Doubt にかかわる構文の歴史的変化について(2) ——Early Modern English Prose Selections の分析から、九大英文学、査読有、54、2012 年、119－134

〔学会発表〕（計　3 件）

① 家入　葉子、The *to*-infinitival Construction of the Verb *convince* in Contemporary American English、4th International Conference on the Linguistics of Contemporary English、2011 年 7 月 20 日、オスナブルック

② 家入　葉子、Adverbial Clauses in the *Paston Letters*、17th International Conference on English Historical

Linguistics、2012 年 8 月 22 日、チューリッヒ

③ 家入　葉子、The Verb *prohibit* and its Complementation in Early Modern English: A Historical Survey、2014 年 1 月 11 日、ホノルル

〔図書〕（計　4 件）

① 家入　葉子　他（共著）、開拓社、『ことばとこころの探求』2012、363－376

② 家入　葉子　他（共著）、John Benjamins、*Middle and Modern English Corpus Linguistics: A Multi-dimensional Approach*、2012 年、59－73

③ 家入　葉子　他（共著）、John Benjamins、*Meaning in the History of English: Words and Texts in Context*、2013 年、211－229

④ 家入　葉子　他（共著）、Osaka Books、*Studies in Middle and Modern English: Historical Change*、2014 年、29－47

〔産業財産権〕

○出願状況（計　0 件）

○取得状況（計　0 件）

６．研究組織

(1)研究代表者

家入　葉子（IYEIRI, Yoko）
京都大学・大学院文学研究科・教授
研究者番号：20264830