様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 １７ 日現在

機関番号：３２６２７
研究種目：基盤研究(C)（一般）
研究期間：2017～2019
課題番号：１７Ｋ０２７８６

研究課題名（和文）英学資料のテキストデータ化に関する研究

研究課題名（英文）Study about converting English-Japanese conversation books into electric data.

研究代表者
常盤　智子（Tokiwa, Tomoko）

白百合女子大学・文学部・教授

研究者番号：６０３６１５５７

交付決定額（研究期間全体）：（直接経費）　　1,500,000円

研究成果の概要（和文）：本研究は明治時代の外国人の手になる口語資料に注目をして、英学資料の電子データ化に取り組んだ。具体的な作業としては、明治6年刊E.M.サトウ著『Kuaiwa Hen』の和文テキスト『春秋雑誌会話篇』についての電子データを作成した後、国立国語研究所のコーパス検索アプリケーション「中納言」での検索を前提としたデータ登録作業を行った。
このほか、期間中に当該資料と並行して、同時期の英学資料2種についても電子データ化の準備に着手することができた。研究期間中に関連の研究発表3件、論文2件の発表を行った。

研究成果の学術的意義や社会的意義
本研究の学術的意義は、明治時代の日本語資料である英学資料を、タグ（コーパスなどの大規模データを検索する際に参照され得るデータ）付きテキストデータへ電子化したことにある。このことは、資料研究の進展に寄与するとともに、電子化が進む日本語史研究の中で外国人による日本語研究の利用価値を示す一助となると考えている。
本研究の社会的意義は、今後の公開作業が進むことで、研究者だけでなく、大学生や大学院生、広く一般の人へもこのデータが利用可能になることがあげられる。当該資料の価値が広く知られる機会となることが期待される。

研究成果の概要（英文）：This study aimed to convert textbooks, written during the second half of the 19th century by Westerners learning Japanese, into electronic data. Ernest Mason Satow's “Shunju Zasshi Kuaiwa Hen, written in 1873, was converted into Extensible Markup Language (XML). Upon completion, the data were registered with Dainagon, searchable through Chunagon, at the National Institute for Japanese Language and Linguistics.
During the study, I presented the research results at three workshops and published two papers in scientific journals.

研究分野：日本語史

キーワード：日本語史　英学資料

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

(1)英学資料研究という面から

英学資料研究が日本語史研究の資料として本格的に活用されるようになるのは、1960 年代ごろからのことである。松村明、森岡健二、古田東朔らをはじめとして、近年では、金子弘「幕末明治期日本語学関連電子化資料」http://home.soka.ac.jp/~ hkaneko/etxt/、大久保恵子〔1999〕『チェンバレン『日本語口語入門』第２版　翻訳 付索引』笠間書院、櫻井豪人〔2009-2013〕「アーネスト・サトウ 『会話篇』 part 2 訳注稿(1)～(7)、補遺」『茨城大学人文学部紀要 人文コミュニケーション学科論集』7~ 14、木村一〔2015〕『和英語林集成の研究』明治書院、申請者〔2015〕『英学会話書の研究』武蔵野書院、等によって、資料研究が行われつつある状況であった。

そもそも、英学資料は日本語史研究において、国内資料にはない特徴を持つ重要な資料群であるが、研究資料は膨大であるわりに、実際に本格的に活用されている資料は限られているのが現状であった。また、比較的よく知られ、用いられることの多い資料にも、研究の余地があると考えられた。

(2)コーパス研究という面から

他方、研究開始時において、日本語コーパスの活用が進められていた（現在も進行中)。

国立国語研究所・言語資源研究系の共同研究プロジェクトとして 2009 年に開始された「通時コーパスの設計」による成果『日本語歴史コーパス』が公開され、『日本語学』（明治書院）臨時増刊号（33-14）や、近藤泰弘・田中牧郎・小木曽智信〔2015〕『コーパスと日本語史研究』などが続々と出版されていた時期であった。

この時期、通時コーパスはすでに日本語史研究の大きな流れの一つとなっていたが、国立国語研究所で公開されている『日本語歴史コーパス』には英学資料のデータは収録されていなかった。折しも申請者自身が「通時コーパスの構築と日本語史研究の新展開」に共同研究員として、参加する機会を得た時期でもあったため、日本語史研究資料の一つとして英学資料を活用するための基礎研究を行うことを計画した。

２．研究の目的

本研究は日本語史研究において、英学資料のデータ化を通して英学資料を日本語史研究の研究資料として広く有効に活用させることを目的とした。

具体的には、幕末明治初期における会話書を中心に、ローマ字で表記された日本語資料の利点を生かしつつ、データ化を行うための諸問題について研究を行うこと、また、研究期間中に、英学資料の特定の資料に焦点を絞り、公開できるデータを完成させることに取り組んだ。

ローマ字で表記された資料のデータ化という作業を通して、新しい知見が得られ、今後の日本語研究への新しい契機となること、また、申請者が継続している英学資料（特に会話書）に関する研究の重要な基盤となることを期待したものであった。

３．研究の方法

日本語史資料としての資料研究を行った上で、選択された資料について、国立国語研究所で公開の進む資料との一括利用を想定し、XML(Extensible Markup Language)データの作成を行った。試作を重ねつつ、タグ（メタデータ：コーパスなどの大規模データの検索等に参照され得るデータ）を付ける作業を行うという方法を取った。XML データ作成の大枠については、近藤明日子〔2014〕「『国民の友コーパス』解説書第 1.1 版」国立国語研究所、を参照し作業を進めた。

なお、資料の選定には、すでに先行研究として、金子弘氏により、「幕末明治期日本語学関連電子化資料」http://home.soka.ac.jp/~ hkaneko/etxt/として、重要資料のいくつかについて、テキストデータでの公開が行われていたため、そちらとは重ならないような資料や扱い方を選定すること、また、テキストデータに加えて XML データを作成することとした。

４．研究成果

(1)概要

研究期間全体の成果は、対象資料に E. M. サトウ著『Kuaiwa Hen』（明治 6 年刊）の国字テキスト『春秋雑誌会話篇』を対象とすることに決定し、データ化を一通り終えたということである。

当該資料は国字テキストではあるものの、同内容のローマ字日本語テキスト、対訳英文、解説書が附属した多角的検討が可能な資料である。また、検討過程において、日本語表記の学習書としても、非常に工夫の凝らされたものであることも判明した。当該資料を選定したことにより、XML データ作成時のタグ付け作業において、対訳英文やテキスト附属の解説書を援用

することにより言語分析ができたという事例もあり、新しいアプローチができたものと考えている。研究期間を通して、関連の研究発表 3 件と研究論文 2 件を公開した。以下、年次ごとに研究の経過と成果をまとめる。

(2)初年度の経過と成果

まず、初年度は資料選定とデータ化に際しての課題の抽出を行った。

資料の選定は上述の通り、『春秋雑誌会話篇』を対象とすることに決定した。その理由は、上述の通り資料形態の特性によるものであった。

データ化に関する作業としては、作業協力者 1 名とともに版本からの翻字作業を含めて、第一次データ入力・点検を終了した。

また、その後のデータへのタグ付け等の検討・調整に向けて、ローマ字資料のデータ化の問題について、国立国語研究所のコーパス設計関係者からヒアリングを行い、理論的な枠組みの理解を行った。

さらに、データ化の作業と並行して、英学資料・対象資料の特徴を生かしながらデータ化を進めるための課題について検討を行い、2 件の口頭発表の機会を得た（「英学資料の可能性―会話書を中心に―」通時コーパスの構築と日本語史研究の新展開　コーパス活用班　近世・近代グループ・文体・資料性グループ合同研究会　2017 年 8 月 20 日、「英学資料とコーパス―『会話篇』の試み―」通時コーパスシンポジウム 2018　2018 年 3 月 10 日）。

当該年度の課題として、仮名遣いに相当するローマ字遣いの補正や、間違った反映も含めた歴史的仮名遣いの反映の補正をどのように処理することが適切か、対訳英文との相互参照をどのように処理することが適切かなどの問題を認識することとなった。

(3)第 2 年目の経過と成果

第 2 年目は、選定した資料のデータ入力、どういった階層構造でデータを作成するかの作業に取り組んだ。作成したデータに対して、タグを付ける作業を行った。

作業は、作業協力者とともにテキストデータに対して、どのタグをどのようにつけていくのかという検討・調整を行いつつ進めた。データ整備、タグ付けの検討・試作に際しては、国立国語研究所の近藤明日子氏に協力を依頼し、国立国語研究所で公開されている日本語歴史コーパスにおけるタグ付けの詳細を学びつつ、作業ツールに関する情報提供や指導をいただく機会を得た。年度末には一通りのタグ付けを終えたものの、なお不明の点や、不統一の面が残った。

上記の作業と並行し、英学資料研究の成果の一つとして「幕末明治期における日英対訳会話書の日本語―数量の多さを表す句との対応から―」（『日本語の研究』第 14 巻 2 号）を公表した。

(4)第 3 年目の経過と成果

第 3 年目は前年度から作業を進めているデータ化を実用化させる作業を行った。第 2 年目に引き続き、近藤明日子氏のご協力を得て、当該データを国立国語研究所のコーパス検索アプリケーション「中納言」の仕様に合わせた形で、短単位という単位ごとに登録する作業を行った。データ整備、点検をすすめ、年度内に一通りの作業を終えることができた。

今後新しい形式（多重形態論情報による本文処理）でのデータ公開に向けて追加の作業を行う予定のため、今年度の公開を視野に入れ引き続き作業を行う予定である。

また、この作業と並行して、作業協力者とともに、同時期の外国人による対訳会話書 2 種（S. R. ブラウン著『Colloquial Japanese』（1863 年刊）、J. F. ラウダー著『Conversations in Japanese and English』（1867 年刊））のデータ化の作業にも着手した。ローマ字テキストに加え、和文テキストが存在していても誤植が散見される前者や、和文テキストの全くないタイプの後者のデータ化については、取り組むべき課題も多く、完成には至っていないが、今後の研究への足掛かりを得ることができた。

このほか、資料探求として、東北大学附属図書館での資料閲覧を行った。

当該年度の研究成果の発表としては、関連の研究発表 1 件「ローマ字資料を用いたコーパス作成における諸問題」（通時コーパスの構築と日本語史研究の新展開　通時コーパス活用班　近世グループ　第 1 回研究発表会　2019 年 12 月 22 日）と論文 1 件「日本語表記の学習書としての『春秋雑誌会話篇』」（『近代語研究』第 21 集　武蔵野書院 9 月）の公表を行った。

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　0件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 常盤智子	４．巻 14-2
２．論文標題 幕末明治期における日英対訳会話書の日本語 ――数量の多さを表す句との対応から――	５．発行年 2018年
３．雑誌名 日本語の研究	６．最初と最後の頁 18-33
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.20666/nihongonokenkyu.14.2_18	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 常盤智子	４．巻 21
２．論文標題 日本語表記の学習書としての『春秋雑誌会話篇』	５．発行年 2019年
３．雑誌名 近代語研究（978-4-8386-0723-5）	６．最初と最後の頁 43-62
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 常盤智子
２．発表標題 英学資料の可能性―会話書を中心に―
３．学会等名 通時コーパスの構築と日本語史研究の新展開　コーパス活用班　近世・近代グループ・文体・資料性グループ合同研究会
４．発表年 2017年

１．発表者名 常盤智子
２．発表標題 英学資料とコーパス―『会話篇』の試み―
３．学会等名 通時コーパスシンポジウム
４．発表年 2018年

１．発表者名 常盤智子
２．発表標題 ローマ字資料を用いたコーパス作成における諸問題
３．学会等名 通時コーパスの構築と日本語史研究の新展開　研究発表会　（通時コーパス活用班　近世グループ　第1回研究発表会）
４．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究協力者	近藤　明日子 (KONDO ASUKO)		