

令和 2 年 6 月 16 日現在

機関番号：32414

研究種目：若手研究(B)

研究期間：2016～2019

課題番号：16K16839

研究課題名(和文)中国語における形態素解析システムの構築と中国語基本語彙研究に関する試み

研究課題名(英文) Construction of morphological analysis system in Japanese and trial on basic vocabulary in Chinese

研究代表者

氷野 善寛 (HINO, Yoshihiro)

目白大学・外国語学部・専任講師

研究者番号：80512706

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究は、明治初期に北京官話を対象とした中国語教育が始まってから現代までの中国語教育史、そしてその中で教えられてきた中国語の語彙がどんなものであったのかということ、中国語教材という視点から検証できないかと考え、様々な視点から原資料に対する検証を試みたり、教科書が収録する単語を解析するためのシステム構築を行ったりしながら、中国語教材を中心としたデータベースの構築を行ったものである。研究成果の一部である、形態素解析システムと中国語教材の書誌データベース及び、中国語基本語彙検証データベースについてはウェブ上で公開している。

研究成果の学術的意義や社会的意義

北京官話を対象とした中国語教育は明治期から現在まで100年以上もの間、実施されてきている。歴史を紐解き、それぞれの時代で学ばれてきた中国語がどういったものだったのかを知ることは、単に教材の歴史を知るだけではなく、それぞれの時代の人々がどういった眼差しで中国を見ていたのかということまでも知ることができる。しかし会話書、教科書といった資料は図書館などの公の機関の蔵書になりづらく、永らく研究対象にもならず、研究を実施するための基礎資料も限定されたものしかなかった。そこで本研究ではそういった資料を積極的に収集、データベース化することで、今後当該分野の研究環境を整えるという意義があったと考える。

研究成果の概要(英文)：This study is about the history of Chinese language education from the beginning of the Chinese language education for Beijing scholarship in the early Meiji era to the present day, and what was the vocabulary of Chinese taught in it. I tried to verify the source material from various viewpoints and tried to verify the source material from various viewpoints, and while constructing a system to analyze the words contained in the textbook, focusing on the Chinese teaching material. This is a database construction.

A morphological analysis system, a bibliographic database of Chinese teaching materials, and a Chinese basic vocabulary verification database, which are part of the research results, are available on the Web.

研究分野：中国語教育

キーワード：中国語教育史 基本語彙

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

本研究課題『中国語における形態素解析システムの構築と中国語基本語彙研究に関する試み』は『通時的中国語基礎語彙データベースの構築と語彙教育に関する研究』(平成24年~26年、若手研究(B)、研究課題番号:24720194)を発展的に継承するものである。

先の研究課題では主に明治期以降の中国語教材の収集や目録編集、その中で行われてきた語彙教育に関する考察を行った。本研究ではさらに、各教材が扱う語彙を量的に調査することにより、中国語教育における学習基本語彙の変遷をより具体的に明らかにし、さらに現在の中国語教育における学習語彙の研究へとつなげていきたいと考えている。現在、大学・高校・民間など多くの教育機関で中国語教育が行われ、毎年数多くの教材が出版されているが、これが「典型」という教材が見当たらない。それぞれの教材は、どこで学ぶのか、誰を対象としているのか、どこで使う中国語を教えようとしているのかなどの条件に従って、各教育機関の授業のコマ数や授業計画によりその分量、内容、学習順序などが異なる。語彙学習一点について取り上げても教員や学習者はどの教材を使えばどういった内容を教えられるのか、あるいは学べるのかといったことを把握するのは学習語彙の範囲や数量について統一された明確な基準がないこともあり容易ではない。学習語彙の把握は非常に難しく、各教材で学べる語彙の範囲や傾向などについて容易に知ることができない。また中国語教材の中で扱われる文章(=テキスト)は時代の流れや社会の変化とともに変わり、採用される語彙や文体なども少しずつ変化しており、申請者はこのような変化が学習語彙に及ぼす影響の研究が不可欠であると考えている。この変化を150年という長い期間で捉えることにより中国語の学習語彙がどのように変化してきたかという根本的な問題を捉えることができると考えている。

量的な語彙分析を行うためには、コンピュータを活用した形態素解析が必要不可欠であるが、現在日本国内はもとより世界中を探しても自由に利用でき、かつ長期間の語彙の変化を通時的に分析できる中国語を対象とした形態素解析システムは存在しない。ここで述べる形態素解析とは対象言語の文法構造や辞品詞等の情報付きの単語リストを情報源として用い、言語で意味を持つ最小単位に分割し分析する作業を指す。そこで本研究ではまず中国語教材が扱うテキストデータをベースに形態素解析用の辞書を作成し、その辞書を扱える自前のシステムを開発することからはじめようと考えた。このシステムでは、単に形態素解析ができるだけではなく、教材ごとの学習語彙数や語彙の傾向といった詳細な情報を「可視化」できるようにしたいと考えている。次に対象とする教材のメタ情報を扱うデータベースも必要である。中国では、中国民族大学や中山大学には2000年以降に刊行された中国語教材を対象としたデータベースが存在し、教授対象、レベル、編集者などの情報を集めている。しかしながらこれらのデータベースは内部資料としてのみ利用されており、オープンな環境で利用提供されておらず、各大学内の一部の研究者が教材研究目的でのみ利用しているのが現状である。日本国内の教材についても教材のメタ情報の収集とデータベース化も必須の作業であり、本研究を通じてそれらを構築したい。そして構築するデータベースやシステムは原則として全てオープンアクセス化したいと考えている。

2. 研究の目的

北京官話を対象とした中国語教育は明治期から現在まで100年以上もの間、実施されてきている。歴史を紐解き、それぞれの時代で学ばれてきた中国語がどういったものだったのかを知ることが、単に教材の歴史を知るだけではなく、それぞれの時代の人々がどういった眼差しで中国を見ていたのかということまでも知ることができる。これは単に過去の歴史を知るというだけではなく、現代において中国語教育がどうあるべきかということを考えることにも直結すると考えている。しかし会話書、教科書といった資料は図書館などの公の機関の蔵書になりづらく、永らく研究対象にもなっておらず、研究を実施するための基礎資料も限定されたものしかなかった。そこで本研究ではそういった資料を積極的に収集、データベース化することで、この分野の研究を資料面やデータ面から支えることが一つの目的である。

また明治初期に北京官話を対象とした中国語教育が始まってから現代までの中国語教育史、そしてその中で教えられてきた中国語の語彙がどんなものであったのかということ、中国語教材という視点から検証できないかと考え、様々な視点から原資料に対する検証を試みたり、教科書が収録する単語を解析するためのシステム構築を行ったりしながら、中国語教材を対象としたデータベースを構築することがもう一つの目的である。

さらに、この数年、中国の世界漢語教育史研究会を中心に過去の世界中の中国語教育や中国語教材に対する研究が加速してきている、本研究を進めることは、こういった研究活動の発展に寄与することにつながると考えている。

3. 研究の方法

中国語解析システムの版をベースに通時的・共時的な解析ができるシステムを構築する。

・デジタル化・テキスト化する中国語教材を選定し、タグ付け作業などを通じて解析用辞書作成の基礎データを作る。作成したデータをもとに、形態素解析用の辞書を構築する。

・構築した形態素解析辞書は、構築しつつある形態素解析システムを利用して次回以降の中国語の分析に用いる。この作業を繰り返し、中国語形態素解析辞書を拡充していく。次にメタ情報を提供する中国語教材データベースとの連携を考える。作成した形態素解析システムを利用して

通時・共時双方から基本語彙の変遷や状況を研究する。研究成果については国内外の学会で随時発表を行い、必要に応じて論文にまとめ、学術誌に発表する手法をとった。

4. 研究成果

(1) 『中国語基本語彙検証データベース』

<https://www.ch-analysis.org/>

1870-2018年に日本で刊行された中国語の会話書、教材を中心とした書誌を集めた「教材データベース」、同期間の会話書、教材から任意の書籍を選び、全文データをコーパスとして収録した例文データベース、その例文に対して形態素解析を実施し、抽出した単語を教材データベースに登録される刊行年のデータと紐づけることで実現した「中国語基本語彙検証データベース」からなる複合的なデータベースである。今回の研究期間では全ての教材をデジタル化し、登録することは研究期間と予算の範囲の関係でできなかったが、データの登録作業は科研費の研究期間が終わった現在も継続的に続けており、今後のさらなるデータベースの充実や同分野の研究への応用が期待できる。

(2) 『Chinese Text Analyzer』

<https://www.ch-staion.org/cta/>

『Chinese Text Analyzer』は中国語の形態素解析のために構築した Mecab をベースに利用したシステムである。『中国語基本語彙検証データベース』にテキストデータを登録する際の処理に利用するために、既存のシステムを改良する形で、上記データベースに組み込んでいる。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 水野善寛	4. 巻 第18号
2. 論文標題 形態素解析の中国語教育への応用	5. 発行年 2018年
3. 雑誌名 漢字文献情報処理研究	6. 最初と最後の頁 pp.51-55
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 -
1. 著者名 水野善寛	4. 巻 なし
2. 論文標題 鱒澤彰夫氏寄贈図書の構成と稀覯本	5. 発行年 2017年
3. 雑誌名 関西大学東西学術研究所蔵鱒澤文庫目録	6. 最初と最後の頁 26-42
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 -
1. 著者名 水野善寛・齊燦	4. 巻 なし
2. 論文標題 『北京官話全編』語彙索引	5. 発行年 2018年
3. 雑誌名 『北京官話全編の研究』	6. 最初と最後の頁 175-954
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 -
1. 著者名 水野善寛	4. 巻 49
2. 論文標題 19～20世紀の中国語教育史を研究するための資料	5. 発行年 2016年
3. 雑誌名 東西学術研究所紀要	6. 最初と最後の頁 515-527
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 有
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 水野善寛	4. 巻 53
2. 論文標題 昭和初期の子供向けの中国語教材の一端 めんこ・かるた・新聞	5. 発行年 2020年
3. 雑誌名 東西学術研究所紀要	6. 最初と最後の頁 209-229
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 有
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 -

〔学会発表〕 計9件(うち招待講演 0件/うち国際学会 5件)

1. 発表者名 水野善寛
2. 発表標題 中国語教育における中国語の分割システム
3. 学会等名 FLEXICT Expo 2018
4. 発表年 2018年

1. 発表者名 水野善寛
2. 発表標題 中国語の形態素解析
3. 学会等名 東西学術研究所・関西大学アジア・オープン・リサーチセンター合同ワークショップ「日中自然言語の処理に関して」
4. 発表年 2018年

1. 発表者名 水野善寛
2. 発表標題 日本漢語教材的歴史演变
3. 学会等名 世界漢語教育史研究学会第十屆年会「デジタル化時代におけるグローバル中国語教育史国際シンポジウム(国際学会)」
4. 発表年 2018年

1. 発表者名 水野善寛
2. 発表標題 明治時代以后的日本漢語教材分析
3. 学会等名 漢語教材史国際学術研討会第八屆年会（国際学会）
4. 発表年 2016年

1. 発表者名 水野善寛
2. 発表標題 ウェブで利用できる中国語形態素解析システムの開発 複数の解析用辞書切替による研究・教学での展開
3. 学会等名 日本中国語学会第66回全国大会
4. 発表年 2016年

1. 発表者名 水野善寛
2. 発表標題 『北京官話全編』の語彙学習について
3. 学会等名 東西学術研究所第17回研究例会
4. 発表年 2017年

1. 発表者名 Yoshihiro Hino
2. 発表標題 Morphological Analysis and lexical indexing of Chinese
3. 学会等名 Japanese Association for Digital Humanities, JADH2019（国際学会）
4. 発表年 2019年

1. 発表者名 水野善寛
2. 発表標題 基本語彙の変遷を検証するための通時型中国語教材コーパスの構築について
3. 学会等名 日本中国語学会第69回全国大会（国際学会）
4. 発表年 2019年

1. 発表者名 水野善寛
2. 発表標題 関于日本漢語課本語料庫の初歩設計
3. 学会等名 世界漢語教育史研究学会（国際学会）
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 内田慶市・吾妻重二・原田正俊・篠原啓方・水野善寛	4. 発行年 2017年
2. 出版社 関西大学東西学術研究所	5. 総ページ数 678
3. 書名 関西大学東西学術研究所蔵鱒澤文庫目録	

〔産業財産権〕

〔その他〕

<p>『中国語基本語彙検証データベース』 https://www.ch-analysis.org/ 『Chinese Text Analyzer』 https://www.ch-staion.org/cta/</p>

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
--	---------------------------	-----------------------	----