

京都大学人文科学研究所共同研究実績・活動報告書

(5年計画の 4年度目)

1. 研究課題

(和文) 東アジア古典文献コーパスの研究

(英文) Study on Corpus of Ancient Chinese

2. 研究代表者

(氏名) 安岡孝一

3. 研究期間

平成20年 4月 から 平成25年 3月 まで

4. 研究目的 (400字程度)

わが国の文化的活動は、奈良時代より随時もたらされた大量の中国の漢文と、それを訓読という手法によって解釈した歴史なしには語るができない。特に、近世、明治～昭和初期における漢文訓読という営為は、伝統的な日本の思想基盤を見直す作業であったとともに、近代国家としての思想基盤を再構築する過程において、重要な位置を占めるものであった。

本共同研究の構想は、その前身となった共同研究班「漢字情報学の構築」(2004年4月～2008年3月、班長：安岡孝一)で培われた知識、特に、コンピュータを用いた漢文解析に関する研究の中から生まれた。その中で、われわれは、日本における訓読が、元となる漢文とその読み下し文との間でそれぞれ異なる文法構造を橋渡しするための手法であることから、漢文の文法解析において有効に働きうるはずだ、という感触を得た。すなわち、訓読という手法を情報学的な視点から再検討し、訓読漢文コーパスともいべきものを作成して、それを他の一般的な漢文(白文)に適用することで、漢文の意味構造を解析可能になるだろう、という予想である。

本共同研究では、明治～昭和初期に日本国内で作成された訓読漢文テキストをコーパス化し、それを基に、漢文の意味構造を解析するシステムの研究・開発をおこなっている。すなわち、これまで漢文を読むための技法に過ぎなかった訓読を、コンピュータによる文法解析メソッドの一つとして、情報学的視点から捉えなおしている。この結果、文法的な構造化がおこなわれずに単なる文字列のままに放置されている大量の漢文に対して、その意味構造を解析することが可能となり、漢文の内容理解への大きな足がかりになると考えられる。

5. 本年度の研究実施状況 (400字程度)

平成22年度は、漢文コーパスの制作作業の準備、および制作環境の構築を中心に、共同研究をおこなった。なお、本研究班では、参加者全員が文献や書籍を見ながら論じ合うという

スタイルを取っているため、特定の発表者等は記さないことにする。

2011年4月15日

科学研究費補助金の配分と今後の研究計画について

2011年5月20日

漢文大系PDF

Gitの新規ディレクトリ製作

2011年6月3日

Emacs For Mac OS X

反切に関するコーパスデータ検討

2011年6月17日

Emacs23 (Cocoa Emacs) 入門から中毒まで (IMEパッチの適用)

2011年7月1日

OSDH2011投稿原稿再読み合わせ

複数のEmacs環境の並立

2011年9月16日

Mecab漢文環境の作り方together

2011年10月14日

『日本語通時コーパスの設計について』

『通時コーパスの構築に向けた古文用形態素解析辞書の開発』

2011年11月4日

孟子定本のコーパス作業データ検討

2011年11月18日

漢文corpusのWIKI風インターフェースに関する議論

2011年12月2日

論語集説のコーパス作業データ検討

2012年1月6日

品詞一覧の検討

2012年1月20日

分類語彙表による分類[名詞]

実詞と虚詞

2012年2月10日

分類語彙表による分類[動詞]

2012年3月9日

来年度作業の指針

6. 研究成果の概要 (400字程度)

平成20年度に開始した本研究班は、平成22年度に至り、全国共同利用・共同研究拠点「人文学諸領域の複合的共同研究国際拠点」の一翼を担うこととなった。また、科学研究費補助金(基盤研究(B))「形態素解析のための品詞情報つき古典文献コーパスの構築」(平成22～24年

度)を獲得し、これまで手弁当でおこなってきた漢文コーパス構築作業を、外部資金によって進めることができるようになった。

平成23年度は、『漢文大系』(富山房)の漢文コーパス構築作業を実施するとともに、それにとまなう品詞等の精細化をおこなった。また、これらの構築作業中に発生した種々の技術的問題点に対し、コーパス構築環境そのものの見直しをもおこなった。

7. 共同研究会に関連した公表実績（出版、公開シンポジウム、学会分科会、電子媒体など）

平成20～22年度の本研究の研究成果を、2011年9月12～14日に開催された国際シンポジウム『Osaka Symposium on Digital Humanities 2011』にて発表した。発表タイトル3件は以下の通り。

- ・ Koichi Yasuoka: Toward a Syntactic Analysis of Classical Chinese Texts
- ・ Naoki Yamazaki: Toward Syntactic Frame Retrieval of Classical Chinese Rhymes using Japanese `kun` readings and Syntactic parallelism of couplets
- ・ Tomohiko Morioka: A Prototype of a Classical Chinese Morphological Analyzer based on MeCab

また、本研究のコーパス構築環境に関して、2012年3月16日に開催された『東洋学へのコンピュータ利用』第23回研究セミナーで発表した。発表タイトルは以下の通り。

- ・ 守岡知彦: 古典中国語形態素コーパス編集システムの開発

8. 本年度の共同利用・共同研究の参加状況

| 区分 | 所属機関数 | 参加人数 | 延べ人数 |
|------------|-------|-------|-------|
| 学内 | 1 | 4 | 44 |
| 国立大学 | 1 | 1 | 3 |
| 公立大学 | 0 | 0 | 0 |
| 私立大学 | 2 | 3 | 21 |
| 大学共同利用機関法人 | 0 | 0 | 0 |
| 民間・独立行政法人等 | 0 | 0 | 0 |
| 外国の研究機関 | 0 | 0 | 0 |
| (うち大学院生) | (0) | (0) | (0) |
| 計 | 4 | 8 | 68 |

※当該年度の共同利用・共同研究参加者の所属機関数、参加人数、延べ人数を区分に応じて記入して下さい。

※「学内」の所属機関数は「学部数」等を記入して下さい。

※参加人数及び延べ人数の算出方法は、以下の例に基づき算出して下さい。

(例)

- ・ 1つの共同利用・共同研究課題で2人を共同研究員として3日間受け入れた（参加した場合）：参加人数2人、延べ人数6人

9. 本年度 共同利用・共同研究を活用して発表された論文数

| | |
|---------------------|---|
| 論文数 | 4 |
| 上記のうち国際学術誌に掲載された論文数 | 3 |

※研究者がファーストオーサーであること。学内の紀要等に発表されたものを除く

なお、高いインパクトファクターを持つ雑誌等に掲載された論文がある場合、その雑誌、掲載論文、そのうち主な論文の詳細等

| 掲載雑誌名等 | 論文名 | 発表者氏名 |
|--------|-----|-------|
| | | |
| | | |