

京都大学人文科学研究所共同研究実績・活動報告書

(4 年計画の 4 年目)

1. 研究課題

東アジア古典文献コーパスの実証研究

Empirical Research on Digital Analysis of Classical Chinese Texts

2. 研究代表者氏名

安岡孝一

Koichi Yasuoka

3. 研究期間

2016 年 04 月 - 2020 年 03 月 (4 年度目)

4. 研究目的

2010 年以來、我々が構築を続けてきた漢文コーパスは、MeCab を用いた形態素解析手法を、漢文処理に適用するものである。この漢文コーパスでは 4 階層の品詞体系を採用しており、その第 2 層は「名詞」「代名詞」「数詞」「動詞」「前置詞」「副詞」「助動詞」「助詞」「感嘆詞」の 9 種類の品詞で構成される。すなわち我々は、従来の漢文文法等で見られた「形容詞」を廃止しているのだが、これが動詞類全体にどのような影響を及ぼしているのかは、必ずしも十分に検討できていない。本共同研究では、漢文コーパスにおける動詞類の実証研究をおこなう。すなわち、実際のコーパスにおいて「動詞」「前置詞」「副詞」「助動詞」の 4 つのふるまいを研究し、さらに下層の意味素性と小素性についても、現在の品詞体系の妥当性を検証する。

In our recent research we have developed a method to analyze classical Chinese texts. In our method, we use our original morphological analyzer based on MeCab. We proposed our original four-level word-class system for classical Chinese on the MeCab-based analyzer. We designed the top level of the word-class system to represent the predicate-object structure of classical Chinese. The second level is the ordinary word-class of classical Chinese, consisting: “noun,” “pronoun,” “numeral,” “verb,” “preposition,” “adverb,” “auxiliary verb,” “particle,” and “interjection.” The third and fourth levels are word-subclasses to describe detailed behavior of the words in classical Chinese texts. In other words, we excluded “adjective” from the second level of our word-class system. But we did not

precisely examine the effect of lack of “adjective” for our morphological analyzer. In this research we will examine the effectiveness of our four-level word-class system, especially examine “verb,” “preposition,” “adverb,” and “auxiliary verb” in the second level. We will also examine the validity of the third and fourth levels of our word-class system.

5. 本年度の研究実施状況

平成 31 年・令和元年度は、『孟子』『論語』『大学』『中庸』の全文を Universal Dependencies で記述した上で、カレル大学の Universal Dependencies 国際プロジェクトに参加し、漢文コーパス全文を WWW 公開した。また、MeCab による漢文形態素解析に加え、UDPipe による依存文法解析ツールを製作し、UD-Kanbun という python モジュールとして公開した。さらに、自動訓読ツール UD-Kundoku も合わせて公開し、我々の漢文品詞体系および Universal Dependencies にもとづく依存文法解析手法が、漢文を「読む」という局面において、実証的に応用可能であることを示した。

6. 研究成果の概要

最終報告書に記載

7. 本年度の研究実施内容

2019-04-12 2019 年度活動方針

2019-04-26 UD_Classical_Chinese-Kyoto (四書版)公開準備

2019-05-11 人文科学とコンピュータ第 120 回研究会『Universal Dependencies の拡張にもとづく古典中国語(漢文)の直接構成鎖解析の試み』

2019-05-24 Universal Dependencies 2.4 公開

2019-06-14 『Classical Chinese Sentence Segmentation』

2019-06-28 UDPipe Visualizer with Immediate Catena Tree [Graphviz 併用版]

2019-07-12 『大學』平行コーパスへの足掛かり

2019-09-13 UD-Kanbun (古典中国語依存文法解析のための python3 モジュール)公開

2019-09-27 漢文 Universal Dependencies からの「文構造」抽出

2019-10-11 『古詩文断句』動作検証

2019-10-25 『一种基于循环神经网络的古文断句方法』

2019-11-08 『大規模格フレームによる解候補削減を用いた ニューラルネットゼロ照応解析』

2019-11-22 Universal Dependencies 2.5 公開

2019-12-14 じんもんこん(・)2019 『漢日英 Universal Dependencies 平行コーパスとその差異』

2019-12-15 UD-Kundoku (古典中国語自動訓読のための python3 モジュール)公開

2020-01-17 共同研究まとめ

2020-03-06 東洋学へのコンピュータ利用

8. 共同研究会に関連した公表実績

山崎直樹「古典中国語のテキストをいかに切り分けるか」(開篇、Vol.37、2019年4月) 安岡孝一「Universal Dependencies の拡張にもとづく古典中国語(漢文)の直接構成鎖解析の試み」(情報処理学会研究報告、Vol.2019-CH-120、No.1、2019年5月) Koichi Yasuoka 「Universal Dependencies Treebank of the Four Books in Classical Chinese」(DADH2019: 10th International Conference of Digital Archives and Digital Humanities、2019年12月) 安岡孝一「漢日英 Universal Dependencies 平行コーパスとその差異」(人文科学とコンピュータシンポジウム「じんもんこん 2019」、2019年12月)

9. 研究班員

所内

Christian Wittern、池田巧、守岡知彦、白須裕之

学内

学外

山崎直樹(関西大学)、二階堂善弘(関西大学)、師茂樹(花園大学)、鈴木慎吾(大阪大学)

10. 共同利用・共同研究の参加状況

区分	機関数	参加人数				延べ人数			
		総計	外国人	大学院生	若手研究者	総計	外国人	大学院生	若手研究者
所内	1	6 (1)	1 (0)	0 (0)	0 (0)	32 (6)	10 (0)	0 (0)	0 (6)
学内	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
国立大学	1	1 (0)	0 (0)	0 (0)	0 (0)	3 (0)	0 (0)	0 (0)	0 (0)
公立大学	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
私立大学	2	3 (0)	0 (0)	0 (0)	0 (0)	12 (0)	0 (0)	0 (0)	0 (0)
大学共同利用機関法人	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

独立行政法人等公的研究機関	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
民間機関	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
外国機関	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
その他	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
計	4	10 (1)	1 (0)	0 (0)	0 (0)	47 (6)	10 (0)	0 (0)	0 (6)

※()内には、女性数を記載

11. 本年度 共同利用・共同研究を活用して発表された論文数

参加研究者がファーストオーサーであるものを対象

総論文数	4(4)
国際学術誌に掲載された論文数	1(1)

※()内には、拠点外の研究者による成果(内数)を記載

12. 費目の 30%を超える大幅な変更があった場合の変更理由

なし

13. 次年度の研究実施計画

なし

14. 研究成果公表計画および今後の展開等

『孟子』『論語』『大學』『中庸』の古典中国語デジタル・コーパスは、カレル大学の **Universal Dependencies** 国際プロジェクトの一部として、公開をおこなった。この成果にもとづき、漢文の自動解析ツール **UD-Kanbun** と、自動訓読ツール **UD-Kundoku** を、python モジュールとして公開した。今後は、新規共同研究班「古典中国語のコーパスの研究」において、さらなる応用と発展をはかる予定である。