京都大学人文科学研究所共同研究最終報告書

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 - 2019 年 03 月

4. 研究目的

2010 年以来、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。

5. 研究成果の概要

古典中国語(漢文)における形態素解析手法を発展させて、さらに文法解析へと展開すべく、数々の手法を検討した。具体的には、Chomsky 流の文法解析手法およびその亜種は古典中国語への適用が難しく、Мельчук流の依存文法(Dependency Grammar)による解析手法が、古典中国語においては非常に有用であることが明らかとなった。この知見にもとづき、現代的な依存文法記法である Universal Dependencies を用いて、『孟子』『論語』『大學』『中庸』の文法記述をおこなって、デジタル・コーパスの形で WWW で公表した。また、『孟子』『論語』『大學』『中庸』等を機械学習した依存文法解析エンジンを制作した上で、大学入試センター試験『国語』の問題のうち、漢文の本文部分に対して、どの程度の自動解析がおこなえるかを検証中である。

6. 共同研究会に関連した公表実績

守岡知彦「古典中国語 UD コーパスの IPFS を用いた表現の試み」(情報処理学会研究報告, Vol.2018-CH-118 (2018 年 8 月), No.6, pp.1-7)　安岡孝一「Universal Dependencies にもとづく古典中国語(漢文)の依存文法解析」(センター研究年報 2018 (2018 年 10 月))　安岡孝一「古典中国語(漢文)の依存文法解析と直接構成素解析」(漢字文献情報処理研究, 第 18 号(2018 年 10 月), pp.56-62)　安岡孝一「漢文の依存文法解析と返り点の関係について」(日本漢字学会第 1 回研究大会予稿集 (2018 年 12 月), pp.33-48)

7. 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、すでに https://corpus.kanji.zinbun.kyoto-u.ac.jp/gitlab/Kanbun/ud-kanbun で公開中であり、『孟子』については冊子(センター研究年報 2018 別冊)の形で出版もおこなった。今後は、古典中国語デジタル・コーパスの拡充をおこなうと同時に、依存文法解析エンジンの解析性能を上げるべく、解析手法についてさらなる研究をおこないたい。