

京都大学人文科学研究所共同研究実績・活動報告書

(3年計画の2年目)

1. 研究課題

古典中国語のコーパスの研究

Study of Classical Chinese Corpora

2. 研究代表者氏名

安岡孝一

Koichi YASUOKA

3. 研究期間

2020年4月-2023年3月(2年目)

4. 研究目的

2010年以来、われわれが構築を続けてきた古典中国語(漢文)コーパスは、MeCabを用いた形態素解析を古典中国語に適用した上で、UDPipeを用いた依存文法解析を適用するものである。これにより、単語の品詞や、単語と単語の係り受け関係を、自動で解析できるようになった。

本共同研究では、古典中国語に対する形態素解析と依存文法解析をさらに押し進め、単語より大きな単位、すなわち句や文について、それらの振る舞いや関係性を解析すべく、さらなる古典中国語解析手法を研究・開発する。

Since 2010, we have developed Classical Chinese Corpora. We first constructed the Corpora using MeCab-Kanbun, a morphological analyzer for Classical Chinese texts. Then we applied UD-Kanbun, a dependency parser based on Universal Dependencies, into the Corpora. Using the Corpora, now we can analyze Classical Chinese texts in word-level: word segmentation (tokenization), Part-Of-Speech tagging, and dependency parsing.

In this study, we will investigate to analyze Classical Chinese texts in phrase- and sentence-levels, enhancing the Classical Chinese Corpora.

5. 本年度の研究実施状況

古典中国語(漢文)Universal Dependenciesを検討しつつ、実際にコーパス化をおこなった。具体的には、詩文として『楚辞』と『唐詩三百首』を、漢訳仏典として『佛説阿彌陀經』『金剛般若波羅蜜經』『維摩詰所説經』『摩訶般若波羅蜜大明呪經』を検討対象とし、順次コーパス化をおこなった。また、これらのコーパスのうち、検討が終了したものから、Universal Dependencies 2.8.1ならびにUniversal Dependencies 2.9として、カレ

ル大学 LINDAT/CLARIN と共同で WWW 公開した。また、これまでの研究成果を、論文『古典中国語（漢文） Universal Dependencies とその応用』としてまとめ、情報処理学会論文誌 2022 年 2 月号に掲載した。

6. 本年度の研究実施内容

- 2021-04-09 『楚辭』の Universal Dependencies 化
- 2021-04-23 『漢文訓読の初期条件(初稿)なぜ孤立語を膠着語に変換できたのか?』
- 2021-05-07 『古漢語音韻データベース「諸家詩經韻読」の構築』
- 2021-05-21 Universal Dependencies 2.8.1
- 2021-06-04 『織田佛教大辞典』の語彙分類
- 2021-07-02 『世界の Universal Dependencies と係り受け解析ツール群』
- 2021-07-16 『佛説阿彌陀經』の Universal Dependencies 化
- 2021-07-30 東洋学へのコンピュータ利用 第 34 回研究セミナー 説文小篆に対する漢字構造記述の試み 発表者 守岡知彦 司会 安岡孝一
- 2021-09-03 『金剛般若波羅蜜經』の Universal Dependencies 化
- 2021-09-17 『維摩詰所説經』の Universal Dependencies 化
- 2021-10-01 『面向多領域先秦典籍的分詞詞性一体化自動標注模型構建』
- 2021-10-15 日本語・現代中国語・古典中国語・タイ語係り受け解析ツール esupar
- 2021-11-05 NINJAL サロン第 228 回「日本語 Universal Dependencies における単語分かち書き」報告
- 2021-11-19 Universal Dependencies 2.9
- 2021-12-03 『日本中世和化漢文における非使役「令」の機能』
- 2021-12-17 じんもんこん:-)2021 報告
- 2022-01-21 『唐詩三百首』の Universal Dependencies 化

7. 共同研究会に関連した公表実績

なし

8. 研究班員

所内

安岡孝一、池田巧、Christian Wittern、守岡知彦、白須裕之

学外

山崎直樹(関西大学外国語学部)、二階堂善弘(関西大学文学部)、師茂樹(花園大学文学部)、鈴木慎吾(大阪大学言語文化研究科)

9. 共同利用・共同研究の参加状況

区分	機関数 (必須)	受入人数					延べ人数				
		総計	海外研究者	若手研究者 (40歳未満)	若手研究者 (35歳以下)	大学院生	総計	海外研究者	若手研究者 (40歳未満)	若手研究者 (35歳以下)	大学院生
学内(法人内)	1	4					64				
国立大学											
公立大学											
私立大学	2	3					41				
大学共同利用機関法人											
独立行政法人等公的研究機関											
民間機関											
外国機関											
その他 ※											
計	3	7 (0)	0 (0)	0 (0)	0 (0)	0 (0)	105 (0)	0 (0)	0 (0)	0 (0)	0 (0)

※「その他」の区分受入がある場合
 具体的な所属等名称を記載：例) 高校教員
 無所属の場合は機関数0とカウントし、この欄の記載不要

2021年7月30日「東洋学へのコンピュータ利用」の参加者30人を除く

10. 本年度 共同利用・共同研究を活用して発表された論文数

	共同利用・共同研究による成果として発表された論文数			
			うち国際学術誌掲載論文数	
①人文研に所属する者のみの論文(単著・共著)	3			
②人文研に所属する者と人文研以外の国内の機関に所属する者の論文(共著)	1			
③人文研以外の国内の機関に所属する者のみの論文(単著・共著)				
④人文研を含む国内の機関に所属する者と国外の機関に所属する者の論文(共著)				
⑤国外の機関に所属する者のみの論文(単著・共著)				

11. 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

12. 次年度の研究実施計画

Universal Dependencies によるコーパスの作成を継続しつつ、四書・詩文・漢訳仏典とは異なる分野にも挑戦したい。具体的には日本漢文の Universal Dependencies を扱いたいのだが、このような研究は計画どおりには進まないことから、現時点では「検討を始める」というところにとどめておく。

13. 次年度の経費

なし

14. 研究成果公表計画および今後の展開等

Universal Dependencies 2.10 およびそれ以後のバージョンを通じて、順次、カレル大学と共同で公開する計画である。