

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3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(3年目)

4. 研究目的

2010年以来、われわれが構築してきた古典中国語(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、新釈漢文大系『日本漢詩』『戦国策』『世説新語』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 2.10 ならびに Universal Dependencies 2.11 として、カレル大学 LINDAT/CLARIN と共同で WWW 公開した。

古典中国語係り受け解析アルゴリズムとしては、従来の Biaffine 法に加え、Question Answering を用いる手法や、goeswith リンクによる手法に挑戦した。特に goeswith リンクによる手法は、古典中国語のみならず、他の孤立語(ベトナム語やタイ語)にも応用可能であり、手法としての広がりが大きそうである。

6. 本年度の研究実施内容

2022-04-15 Universal Dependencies における AUX の扱い
2022-05-20 新釈漢文大系『日本漢詩』
2022-06-03 新釈漢文大系『日本漢詩』Editor 開発
2022-06-17 Transformers の Question Answering を用いた係り受け解析器
2022-07-01 『古漢語詞義注語料庫的構建及应用研究』
2022-07-15 新釈漢文大系『戦国策』
2022-07-29 東洋学へのコンピュータ利用 第 35 回研究セミナー 低品質文字画像を用いた高精細画像からの字形自動再切り出しの試み 発表者 守岡知彦
2022-09-16 nsubj:outer と csubj:outer の追加
2022-09-30 roberta-classical-chinese-base-ud-goeswith による句間リンク抽出
2022-10-21 roberta-classical-chinese-base-ud-goeswith による句間リンク抽出
2022-11-04 roberta-classical-chinese-base-ud-goeswith による句間リンク抽出
2022-11-18 Universal Dependencies 2.11 リリース
2022-12-02 「第 4 回 InDi 學術大會」報告
2022-12-16 roberta-classical-chinese-base-ud-goeswith の解析手法を他の言語に応用する
2023-01-20 共同研究班まとめ

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

鈴木慎吾(大阪大学言語文化研究科)、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)

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
人文研所属 (内女性)		5					64				
京大内 (人文研を除く) (内女性)											
国立大学 (内女性)	1	1					2				
公立大学 (内女性)											
私立大学 (内女性)	2	3					26				
大学共同利用機関法人 (内女性)											
独立行政法人等公的研究機関 (内女性)											
民間機関 (内女性)											
外国機関 (内女性)											
その他 ※ (内女性)											
計	3	9 (0)	0 (0)	0 (0)	0 (0)	0 (0)	92 (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

2022年7月29日「東洋学へのコンピュータ利用」の参加者28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(3)		1(1)	
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(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. 本年度共同利用・共同研究による成果として発行した研究書
なし

12. 博士学位を取得した学生の数(人)

	人数
博士学位を取得した学生の数	0

13. 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

14. 次年度の研究実施計画
なし

15. 次年度の経費
なし

16. 研究成果公表計画および今後の展開等

Universal Dependencies 2.12 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。