

# 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1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(1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

『禮記』の Universal Dependencies コーパスを完成し、つづいて『十八史略』の Universal Dependencies コーパスに着手した。これらのコーパスと、過去に製作した『孟子』『論語』コーパスを合わせ、カレル大学との国際協力により、Universal Dependencies 2.6 (2020

年5月15日リリース)および Universal Dependencies 2.7 (2020年11月15日リリース)として、WWW で公開した。

これらの古典中国語コーパスをもとに、古典中国語形態素解析エンジン「MeCab-Kanbun」および古典中国語係り受け解析エンジン「UD-Kanbun」の解析精度を上げ、PyPI から python3 モジュールとして公開した。また、スタンフォード大学との国際協力により、多言語係り受け解析エンジン「Stanza」に、古典中国語モジュールを実装した。さらに、カレル大学との国際協力により、多言語係り受け解析 API 「UDPipe 2」にも、古典中国語モジュールを実装中である。

## 6. 本年度の研究実施内容

2020-04-10 MeCab-Kanbun と UD-Kanbun

2020-04-24 研究班活動方針

2020-05-08 UD-Chinese の試作

2020-05-22 Universal Dependencies 2.6 リリース

2020-06-05 『禮記』 Universal Dependencies 化完了

2020-06-19 Universal Dependencies Workshop

2020-07-03 『An Advanced Introduction to Semantics: A Meaning-Text Approach』

2020-07-17 Enhanced Universal Dependencies

2020-09-18 『形態素解析部の付け替えによる近代日本語(旧字旧仮名)の係り受け解析』

2020-10-02 CoNLL-U SVG Editor RtoL

2020-10-16 変体漢文の XPOS を UniDic 品詞にする

2020-11-06 『Is POS Tagging Necessary or Even Helpful for Neural Dependency Parsing?』

2020-11-20 Universal Dependencies 2.7 リリース

2020-12-04 『Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection』

2020-12-18 じんもんこん:-)2020 報告

2021-01-15 COMBO-pytorch と UniDic-COMBO

2021-03-05 東洋学へのコンピュータ利用

## 7. 共同研究会に関連した公表実績

Universal Dependencies 2.6 [<http://hdl.handle.net/11234/1-3226>]

Universal Dependencies 2.7 [<http://hdl.handle.net/11234/1-3424>]

## 8. 研究班員

所内

池田巧、ウィッテルン・クリスティアン、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	外国人	若手研究者	若手研究者	大学院生	総計	外国人	若手研究者	若手研究者	大学院生
				(40歳未満)	(35歳以下)				(40歳未満)	(35歳以下)	
学内(法人内)	1	6	1				60	16			
国立大学	1	1					1				
公立大学											
私立大学	2	3					28				
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	4	10	1	0	0	0	89	16	0	0	0
		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

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)	4(4)			
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)				
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

『十八史略』の Universal Dependencies コーパスを完成したい。また、これに合わせ、カレル大学との国際協力により、Universal Dependencies 2.8 (予定)として、何とか WWW 公開したい。ただ、これらのコーパスにより、古典中国語の文法解析の精度が上がるかどうかは、正直なところ五里霧中であり、必要ならば新たな手法の開発と並行して、句や文の範囲を超える文法解析に挑戦したい。

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

カレル大学との国際協力による Universal Dependencies の公開は、今後も継続しておこなう予定である。また、解析エンジンなどは、python3 モジュールであれば PyPI から、それ以外は GitHub から、それぞれ公開する予定である。