

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

古典中国語コーパスの応用研究

Applied Study of Classical Chinese Corpora

2. 研究代表者氏名

安岡 孝一

YASUOKA, Koichi

3. 研究期間

2023年4月-2026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、Conditional Random Fields を用いた形態素解析を古典中国語に適用した上で、Universal Dependencies にもとづく依存文法解析を適用するものである。これにより、古典中国語における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応用して、現代中国語や近代日本語・韓国語・タイ語などの周辺諸言語に対し、文法解析手法を研究・開発する。

Since 2010, we have been developing Classical Chinese Corpora. We first constructed the Corpora using MeCab-Kanbun, a morphological analyzer based on Conditional Random Fields, for Classical Chinese texts. Then we applied SuPar-Kanbun, a dependency parser based on Universal Dependencies, to the Corpora. Using the Corpora and pre-trained language models, we can now analyze Classical Chinese texts by Part-Of-Speech tagging and dependency-parsing.

In this study, we will enhance our method so that it can be applied to other languages, such as Mandarin Chinese, Thai, Japanese, and Korean.

5. 本年度の研究実施状況

『日本書紀』Universal Dependencies の製作に注力し、『日本書紀』の言語構造がどのようなものであるのか、かなりの部分を明らかにした。『日本書紀』では、書写言葉におけるコードスイッチングが、古典中国語(漢文)と上代日本語の間で起こっており、それらの両方が漢字によって記されている。すなわち、漢字の字面だけでは、それがどちらの言語であるかはわからず、読んでみて(すなわち言語処理をおこなってみて)はじめて、どの言語である

かが確定する。結果として、コードスイッチングではあるものの、古典中国語と上代日本語が、実際には入り混じった形となっており、Universal Dependenciesによる記述は何とかできるものの、自動処理は全く出来ない状態である。

6. 本年度の研究実施内容

- 2024.4.19 『日本書紀』の「之」と「者」
- 2024.5.17 『日本書紀』の「之」と「者」
- 2024.6.7 LREC-COLING 2024 文献紹介
- 2024.6.21 『日本書紀』の「之」と「者」と「丹」
- 2024.7.5 『日本書紀』の「之」と「者」
- 2024.7.19 JHPCN 第16回シンポジウム参加報告
- 2024.7.26 『東洋学へのコンピュータ利用』第37回研究セミナー
- 2024.9.6 『日本書紀』の「之」と「者」
- 2024.10.4 『日本書紀』における非漢文
- 2024.10.18 『じんもんこん:-)2024』 ゲラチェック
- 2024.11.15 『日本漢字学会第7回研究大会』 ゲラチェック
- 2024.12.6 『日本書紀』の「之」と「者」
- 2025.1.17 『日本書紀』の「之」と「者」
- 2025.1.31 共通テスト2025国語第5問

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、李媛、劉冠偉

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)、守岡知彦(国文学研究資料館研究部)

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)	1	5	0	0	0	0	52	0	0	0	0						
京大内 (人文研を除く) (内女性)	0	0	0	0	0	0	0	0	0	0	0						
国立大学 (内女性)	1	1	0	0	0	0	2	0	0	0	0						
公立大学 (内女性)	0	0	0	0	0	0	0	0	0	0	0						
私立大学 (内女性)	3	3	0	0	0	0	27	0	0	0	0						
大学共同利用機関法人 (内女性)	1	1	0	0	0	0	13	0	0	0	0						
独立行政法人等公的研究機関 (内女性)	0	0	0	0	0	0	0	0	0	0	0						
民間機関 (内女性)	0	0	0	0	0	0	0	0	0	0	0						
外国機関 (内女性)	0	0	0	0	0	0	0	0	0	0	0						
その他 ※ (内女性)	0	0	0	0	0	0	0	0	0	0	0						
計	6	10	0	0	0	0	94	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	2024年7月26日『東洋学へのコンピュータ利用』の参加者20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		0	
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)	0	(0)	0	(0)
③人文研以外の国内の機関に所属する者ののみの論文(単著・共著)	0		0	
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)	0	(0)	0	(0)
⑤国外の機関に所属する者ののみの論文(単著・共著)	0		0	

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
なし

13. 次年度の研究実施計画

『日本書紀』Universal Dependencies を元にした言語処理システムを構築していきたいが、現時点では「実施計画」を作れる状態ではなく、正直なところ五里霧中である。ただ、上代日本語における古典中国語との混在が、次の時代の日本語の書写体系と言語構造を作ってきたんだろうを考えると、とりあえずは鎌倉・室町期の日本漢文（と呼ばれる文書）に着手することにしている。

14. 研究成果公表計画および今後の展開等

人文情報学創新センターに設置している GitLab サーバーを用いて、Universal Dependencies コーパスの発信を継続する。また、『東洋学へのコンピュータ利用』などの研究セミナーにおいて、経過を順次、報告していきたい。