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(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、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

古典中国語 Universal Dependencies の拡張対象として『日本書紀』を選定し、形態素解析のためのコーパス設計を開始した。『日本書紀』の「漢文」は、大きく α 群と β 群に分かれており、α 群は古典中国語に近いが、β 群は上代日本語の「漢訳」である。実際にコーパスを作りかけてみたところ、β 群の方が明らかに手強い。特に複雑なのが「之」と「者」の用法であり、古典中国語から「かけ離れた」用法が散見される。これらの用法を、Universal

Dependencies においてどう記述するかについて、試行錯誤を繰り返している。

６．本年度の研究実施内容

2023-04-21　過去の共同研究班の成果（抜粋）
2023-05-12 『古事記』と『日本書紀』
2023-06-02　ICBIR 2023 (LDA 2023)報告
2023-06-16 『遍照發揮性靈集』と『法華玄義』
2023-07-21 『日本書紀』Univerfsal Dependencies エディター
2023-07-28 『東洋学へのコンピュータ利用』第 36 回研究セミナー
2023-09-08 『日本書紀』の「之」
2023-09-22 『AI 等の活用を推進する研究データエコシステム構築事業シンポジウム 2023』ゲラチェック
2023-10-06 『日本書紀』の「之」と「者」
2023-10-20 『じんもんこん:-)2023』ゲラチェック
2023-11-17　Universal Dependencies 2.13
2023-12-01 『日本書紀』における非漢文
2024-01-19 『東洋学へのコンピュータ利用』第 37 回研究セミナー
2024-02-02 『日本書紀』の「之」と「者」

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、WITTERN, Christian 、李媛

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 総計	受入人数 海外研究者	受入人数 若手研究者（40歳未満）	受入人数 若手研究者（35歳以下）	受入人数 大学院生	延べ人数 総計	延べ人数 海外研究者	延べ人数 若手研究者（40歳未満）	延べ人数 若手研究者（35歳以下）	延べ人数 大学院生
人文研所属	1	5	0	0	0	0	44	0	0	0	0
（内女性）		(1)	(0)	(0)	(0)	(0)	(4)	(0)	(0)	(0)	(0)
京大内（人文研を除く）	0	0	0	0	0	0	0	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
国立大学	1	1	0	0	0	0	2	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
公立大学	0	0	0	0	0	0	0	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
私立大学	1	3	0	0	0	0	23	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
大学共同利用機関法人	1	1	0	0	0	0	12	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
独立行政法人等公的研究機関	0	0	0	0	0	0	0	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
民間機関	0	0	0	0	0	0	0	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
外国機関	0	0	0	0	0	0	0	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
その他　※	0	0	0	0	0	0	0	0	0	0	0
（内女性）		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
計	4	10	0	0	0	0	81	0	0	0	0
		(1)	(0)	(0)	(0)	(0)	(4)	(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	2023年7月28日『東洋学へのコンピュータ利用』の参加者21人を除く 2024年1月19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）	2		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)	0	(0)
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）	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. 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
『日本書紀』の Universal Dependencies 化を継続する。『日本書紀は』は β 群がかなり手強い、ということがわかってきているので、まずは β 群に注力する。なお、『日本書紀』には「漢文」で書かれた部分以外に、いわゆる「万葉仮名」で書かれた和歌や童謡があるが、これらについても、とりあえず着手だけはしておきたい。

15. 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。