The First International Ancient Chinese Word Segmentation and POS Tagging Bakeoff: Overview of the EvaHan 2022 Evaluation Campaign

Bin Li [1] Yiguo Yuan[1] Jingya Lu[1] Minxuan Feng[1] Chao Xu[1] Weiguang Qu[2] Dongbo Wang[3]

1. School of Chinese Language and Literature, Nanjing Normal University
2. School of Computer and Electronic Information, Nanjing Normal University
3. College of Information Management, Nanjing Agricultural University

E-mail: libin.njnu.at.gmail.com

Abstract

This paper presents the results of the First Ancient Chinese Word Segmentation and POS Tagging Bakeoff (EvaHan), which was held at the Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) 2022, in the context of the 13th Edition of the Language Resources and Evaluation Conference (LREC 2022). We give the motivation for having an international shared contest, as well as the data and tracks. The contest is consisted of two modalities, closed and open. In the closed modality, the participants are only allowed to use the training data, obtained the highest F1 score of 96.03% and 92.05% in word segmentation and POS tagging. In the open modality, the participants can use whatever resource they have, with the highest F1 score of 96.34% and 92.56% in word segmentation and POS tagging. The scores on the blind test dataset decrease around 3 points, which shows that the out-of-vocabulary words still are the bottleneck for lexical analyzers.

Keywords: Evaluation, Ancient Chinese, Word Segmentation, POS Tagging

1. Introduction

EvaHan2022 is the first campaign devoted to the evaluation of Natural Language Processing (NLP) systems for the Ancient Chinese. [1] Unlike English or other western languages, Chinese does not have word boundaries. Thus, word segmentation is a basic task for Chinese language processing. It has received a lot of attention in the literature (Sun and Zou, 2001; Xue et al., 2003). There are five word segmentation bakeoffs for Mandarin Chinese held by SIGHAN (Special Interest Group of Han) workshops during 2003 to 2012 (Sproat and Emerson, 2003; Emerson, 2005; Levow, 2006; Jin and Chen, 2008; Duan et al., 2012) with the highest F1 score around 98% in the open modality test.

Ancient Chinese is a dominant written language during Pre-Qin(before 221BC) and Han dynasties(202BC-220AD). This continued in later dynasties until the 1900s. It is also named as Old Chinese, or Literary Chinese (Wenyan 文言) [2]. There are huge numbers of ancient books written in this language, which requires fast and efficient automatic tools to conduct word segmentation and POS (part-of-speech) tagging. The character, lexicon and grammar of Ancient Chinese differs a lot from the Mandarin Chinese, and the existing Mandarin Chinese lexical analyzers can not run on the Ancient Chinese texts. At the same time, the ancient Chinese has many fewer lexicons and corpora for training and evaluation. Therefore, a standard shared task is needed for developing the Ancient Chinese analyzers.

EvaHan2022 aims to answer two main questions:

- How can we promote the development of resources and language technologies for the Ancient Chinese language?
- How can we foster collaboration among scholars working on Ancient Chinese and attract researchers from different disciplines?

EvaHan2022 is proposed as part of the Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA), co-located with LREC 2022. [3] EvaHan is organized by the Computational Linguistics and Digital Humanities (CLDH) Group at Nanjing Normal University in Nanjing, China. Scorer and detailed guidelines are all available in a dedicated GitHub repository.[4] LT4HALA also holds the shared task for Latin lemmatization and POS tagging (EvaLatin2022), which affords an opportunity for the comparison of the two ancient languages.

2. Task

EvaHan2022 has one joint task, Word Segmentation and POS tagging:

1. Word segmentation is the process of transforming Chinese character sequence to word sequence.
2. POS tagging is the process of labelling the word sequence with its Part-of-Speech identifiers.

In this shared task, a sentence should be automatically parsed from raw text to POS tagged text shown in Table 1. The evaluation toolkit gives the scores on both word segmentation and POS tagging. EvaHan2022 does not accept running results with word segmentation only.

[1] https://circse.github.io/LT4HALA/2022/EvaHan

[2] https://en.wikipedia.org/wiki/Old_Chinese

[3] https://lrec2022.lrec-conf.org/en/

[4] https://github.com/CIRCSE/LT4HALA/blob/master/2022/data_and_doc/

Raw Text with Punctuations	亟請於武公，公弗許。
Annotated Text with word boundaries	亟 請 於 武公 ， 公 弗 許 。
Annotated Text with word boundaries and POS tags	亟/d 請/v 於/p 武公/nr ， /w 公/n 弗/d 許/v 。/w

Table 1: Examples of Word Segmentation and POS Tagging.

3. Dataset

The dataset of EvaHan 2022 is made of texts from the classic historical books *Zuozhuan* (左传) , Shiji (史记) and Zizhitongjian (资治通鉴). The training and gold texts have been automatically punctuated, word segmented and POS tagged, and then manually corrected by Ancient Chinese language experts.

3.1 Data Format

The dataset consists of three parts, a Training dataset and two Test datasets. All the data is distributed following the word segmentation and POS tagging guidelines for Ancient Chinese by Nanjing Normal University (Chen et al. 2013). According to such format, annotations are encoded in UTF-8 plain text files. There are no word boundaries in Chinese texts. Thus, the raw texts contain characters and punctuation. After manual annotation, word boundaries and POS tags are added to the text. As shown in Table 1, each word is labelled with a POS tag, in the form of **Word/POS**. And each word is separated by a space. Punctuations are treated as words too.

3.2 Training Data

The training data contains punctuated, word-segmented and part-of-speech tagged text from *Zuozhuan (左传)*, an ancient Chinese work believed to date from the Warring States Period (475-221 BC). *Zuozhuan* is a commentary on the book *Chunqui (春秋)*, recording the history of the Chinese Spring and Autumn period (770-476 BC).

The files are presented in UTF-8 plain text files using traditional Chinese script. It is released via Linguistic Data Consortium (LDC)[5].

Data Sets	Sources	# Char Tokens	# Word Tokens
Train	*Zuozhuan*	194,995	166,142
Test_A	*Zuozhuan*	33,297	28,131
Test_B	*Shiji, Zizhitongjian*	62,969	55,990

Table 2: Texts distributed as training/test data in EvaHan 2022.

3.3 Test Data

Test data is provided in raw format, with Chinese characters and punctuations. The gold standard test data, which had been manually checked for the evaluation, was provided to the participants after the evaluation.

There are two test datasets. *Test_A* is designed to see how a system performs on the data from a single book. *Test_A* is extracted from *Zuozhuan*, not overlapping with *Train*. *Test_A* has been released by LDC. But the teams are not allowed to use it as training data. There have been several papers reporting their performance on this data (Shi et al., 2010; Cheng 2020 et al., 2020).

Blind *Test_B* is designed to see how a system performs on similar data, texts of similar content but from different books *Shiji (史记)* and *Zizhitongjian (资治通鉴)*. *Test_B* has not been released publicly before EVAHAN. Its size is similar to that of *Test_A*.

4. Evaluation

Each participating team initially had access only to the training data. Later, the unlabeled test data was released. After the assessment, the gold labels for the test data was also released.

4.1 Scoring

The scorer employed for EvaHan is a modified version of the one developed for the SIGHAN2008 (Jin and Chen, 2008). The evaluation aligned the system-produced words to the gold standard ones. Then, Word Segmentation (WS) and Part-of-Speech (POS) tagging were evaluated separately: precision, recall and F1 score are calculated. The final ranking will be based on F1 score.

4.2 Two Modalities

Each participant can submit runs following two modalities. In the *closed* modality, the resources each team could use are limited. Each team can only use the Training data *Train*, and the pretrained model *SIKU-Roberta*[6]. It is the word embeddings pretrained on a very large corpus of traditional Chinese collection, *Siku Quanshu* (四库全书)[7]. Other resources are not allowed in the closed modality.

In the *open* modality, there is no limit on the resources, data and models. Annotated external data, such as the components or Pinyin of the Chinese characters, word embeddings can be employed. But each team has to state all the resources, data and models they use in each system in the final report.

Limits	Closed Modality	Open Modality
Machine learning algorithm	No limit	No limit
Pretrained model	Only SIKU_Roberta	No limit
Training data	Only *Train*	No limit
Features used	Only from *Train*	No limit
Manual correction	Not allowed	Not allowed

Table 3: Limitations on the two modalities.

4.3 Procedures

Training data was released for download from Dec 20, 2021. Test data was released on March 31, 2022, and results were due on 00:00(UTC) April 7, 2022 .

[5] https://catalog.ldc.upenn.edu/LDC2017T14
[6] https://huggingface.co/SIKU-BERT/sikuroberta
[7] https://en.wikipedia.org/wiki/Siku_Quanshu

5. Participants and Results

5.1 Participants

A total of 14 teams took part in the task, submitting 55 running results. Table 4 lists the teams' basic information. Almost all the teams submitted their running results under the closed modality, while only 5 teams attended the open modality. Four files were in wrong formats (marked + in table 4), which have been corrected for evaluation. Two files were submitted overdue (marked * in table 4).

ID	Name	Affiliation	TestA		TestB	
			C	O	C	O
1	BIT1	Beijing Institute of Technology	1	0	1	0
2	BIT2	Beijing Institute of Technology	1	0	1	0
3	BLCU	Beijing Language and Culture University	2	2	2	2
4	BUPT	Beijing University of Posts and Telecommunications	1	0	1	0
5	FDU	Fudan University	2	2	2	2
6	GDUFS	Guangdong University of Foreign Studies	2	0	2	0
7	HIT	Harbin Institute of Technology	2	2	2	2
8	IMUT	Inner Mongol University of Technology	1	0	1	0
9	NJU	Nanjing University	2	0	2	0
10	NJUPT	Nanjing University of Posts and Telecommunications	1*	1^+	1*	0
11	NUAA	Nanjing University of Aeronautics and Astronautics	1	0	0	2
12	THU	Tsinghua University	1	0	1	0
13	ZNNU	Zhongbei College of Nanjing Normal University	1^+	1^+	0	1^+
14	ZYB	Zuoyebang Education Technology (Beijing) Co., Ltd	2	0	2	0
Total files		55	20	8	18	9

Table 4: Participating teams by test datasets and modalities (**C**losed and **O**pen). + files with format correction * submitted overdue

5.2 Results

Table 5-8 list the performances of the teams' systems, sorted by PF (POS tagging F1-score) value (descending). The Precision, Recall and F1 score for Word Segmentation, are shortened as WP, WR and WF. The Precision, Recall and F1 score for Part-of-speech Tagging, are shortened as PP, PR and PF. We categorized the results submitted by the participants as *TestA* Closed, *TestA* Open, *TestB* Closed, and *TestB* Open. The results are ranked by the POS tagging (PF) scores. Most teams participated in closed tests. It can be seen from the four tables that there is a high correlation between word segmentation and POS tagging.

For *TestA*, the highest F1 score of POS tagging is 92.05% in the closed modality. In the open modality, it rises up to 92.56%.

The scores of word segmentation are much higher. FDU scores 96.12% and 96.34% in the closed and open modality. It is remarkble that BUPT scores 96.16% in the closed modality, with a slightly lower score 91.24% for POS Tagging.

For *TestB*, which is designed to see how the systems perform on similar data, the scores all drop down about 3 to 5 points. In the closed modality, FDU achieves 87.77%, only a little lower than 87.87% in the open modality, which means, the outter resources do not help much. The segmentation scores drops to 93.34% and 93.60% in the closed and open modality. The lower performer on *TestB* is possibly caused by the OOV(Out of Vocabulary) words.

ZNNU scores 89.47% in *TestB*, ranking the first place in the open modality. But they did not submit the running file in the closed modality, and this score is even higher than their performance on *TestA*. The outer resources may help them achieve this high score.

Team	**WP**	**WR**	**WF**	**PP**	**PR**	**PF**
FDU	95.39	96.68	**96.03**	91.43	92.67	**92.05**
	95.57	96.67	**96.12**	91.50	92.55	**92.02**
BIT	95.18	96.49	95.83	90.96	92.22	**91.59**
BUPT	95.81	96.52	**96.16**	90.90	91.57	91.24
NUAA	95.63	96.33	95.98	90.88	91.54	91.21
GDUFS	94.85	96.52	95.68	90.34	91.93	91.13
THU	94.70	95.72	95.20	89.59	90.55	90.07
NJU	94.15	95.46	94.80	89.29	90.53	89.90
	94.18	95.44	94.81	89.28	90.47	89.87
GDUFS	92.27	95.46	93.84	88.14	91.18	89.63
BIT2	94.48	94.99	94.74	88.95	89.43	89.19
IMUT	94.67	93.10	93.88	89.73	88.24	88.98
ZYB	94.90	95.07	94.99	88.30	88.46	88.38
ZNNU	92.76	91.45	92.10	88.80	87.54	88.16
ZYB	94.86	94.95	94.90	87.49	87.58	87.53
HIT	90.78	93.03	91.89	84.70	86.80	85.74
	90.81	92.99	91.89	84.72	86.77	85.73
BLCU	91.39	93.22	92.29	84.39	86.09	85.23
	91.39	93.27	92.32	84.20	85.93	85.05
NJUPT*	78.13	86.32	82.03	58.48	64.61	61.39

Table 5 *TestA* closed modality (%)

Team	**WP**	**WR**	**WF**	**PP**	**PR**	**PF**
FDU	95.81	96.88	**96.34**	92.05	93.07	**92.56**
	95.73	96.84	**96.28**	91.88	92.94	**92.41**
ZNNU	92.78	90.18	91.46	88.97	86.48	**87.71**
HIT	91.20	93.49	**92.33**	85.41	87.56	86.47
	91.09	93.41	92.24	85.27	87.45	86.35

BLCU	90.91	92.40	91.65	83.55	84.92	84.23
	90.56	92.29	91.41	83.13	84.72	83.92
NJUPT	78.14	86.31	82.02	57.35	63.35	60.20

Table 6 *TestA* open modality (%)

Team	WP	WR	WF	PP	PR	PF
FDU	94.72	91.99	**93.34**	89.07	86.50	**87.77**
	94.65	91.68	**93.14**	88.98	86.19	**87.57**
BIT	94.48	91.70	93.07	88.40	85.80	**87.08**
GDUFS	94.59	92.70	**93.64**	87.87	86.12	86.99
	92.81	93.20	93.01	86.58	86.94	86.76
BUPT	94.04	90.59	92.28	86.86	83.67	85.24
THU	93.51	90.35	91.90	86.38	83.32	84.82
IMUT	93.65	86.43	89.89	87.05	80.33	83.56
BIT2	93.07	88.90	90.94	85.45	81.61	83.49
ZYB	93.59	89.89	91.70	84.69	81.34	82.98
	93.61	89.97	91.75	84.00	80.74	82.33
NJU	90.00	87.94	88.96	80.89	79.03	79.95
	89.56	87.31	88.42	80.56	78.53	79.53
BLCU	87.72	84.50	86.08	75.32	72.55	73.91
	87.65	84.61	86.10	75.21	72.60	73.88
HIT	82.79	78.82	80.75	71.37	67.95	69.62
	82.19	77.82	79.94	70.21	66.45	68.27
NJUPT*	81.24	85.13	83.14	58.25	61.04	59.62

Table 7 TestB closed modality (%)

Team	WP	WR	WF	PP	PR	PF
ZNNU	95.26	94.79	**95.03**	89.70	89.25	**89.47**
FDU	94.97	92.26	**93.60**	89.16	86.62	**87.87**
	94.81	91.94	**93.35**	88.85	86.16	**87.48**
NUAA	94.50	91.69	93.07	87.79	85.18	86.47
	94.49	91.69	93.07	87.79	85.18	86.46
BLCU	87.09	83.43	85.22	73.99	70.88	72.40
	87.03	83.38	85.16	73.48	70.40	71.91
HIT	83.27	79.30	81.24	71.81	68.38	70.05
	82.23	78.31	80.22	70.77	67.40	69.04

Table 8 *TestB* open modality (%)

5.3 Baselines and Toplines

To provide a basis for comparison, we computed the baseline and possible topline scores for each of the test sets according to the scores in Fourth International Chinese Language Processing Bakeoff (Jin and Chen, 2008).

5.3.1 Word Segmentation

The baseline for ancient Chinese word segmentation is constructed by left-to-right maximal match algorithm using the training set vocabulary. The topline employs the same procedure, but instead uses the test set vocabulary.

Test Set	WP	WR	WF
TestA	84.98	89.20	87.04
TestB	80.43	85.28	82.78

Table 9. Word segmentation baselines (%)

Test Set	WP	WR	WF
TestA	99.04	98.20	98.62
TestB	98.48	97.11	97.79

Table 10. Word segmentation toplines (%)

The word segmentation scores of most teams exceed the baselines in *TestA* and *TestB*. The best scores outperform the baselines by around 10 points as shown in Table 11.

Test set	WP	WR	WF
TestA	+10.83	+7.68	+9.30
TestB	+14.83	+9.51	+12.25

Table 11. The promotion to the baselines of word segmentation (%)

5.3.2 POS tagging

The baseline for ancient Chinese POS tagging is constructed on the test set, word-segmented by the baseline for word segmentation and calculated by generating a list of words and POS tags from the training set. The tagging process is: (1) Tag those words which have only one POS tag in the list; (2) For those words that have not only one tag, the unique most frequent tag in the training set is assigned to them; (3) For each word that does not have a unique most frequent tag, its tag which is the most frequent in the overall training set is assigned to it; (4) Those words that are not in the list are assigned with the most frequent tag in the overall training set. The topline for ancient Chinese POS tagging is constructed on the test sets word-segmented by the topline for word segmentation and calculated by generating a list of words and POS tags from each test set.

The scores of most teams exceed the baselines in *TestA* and *TestB*, as shown in Table 14. And the best POS tagging score exceeds the topline, shown in Table 15.

Test Set	PP	PR	PF
TestA	75.93	79.70	77.77
TestB	66.83	70.87	68.79

Table 12. POS tagging baselines

Test Set	PP	PR	PF
TestA	91.76	90.99	91.37
TestB	89.77	88.51	89.14

Table 13. POS tagging toplines

Test Set	PP	PR	PF
TestA	+16.12	+13.37	+14.79
TestB	+22.87	+18.37	+20.68

Table 14. The promotion to the baselines of POS tagging

Test Set	PP	PR	PF
TestA	+0.29	+2.08	+1.19
TestB	-0.07	+0.74	+0.33

Table 15. The promotion to the topline of POS tagging

5.4 Comparison with EVALATIN

EvaHan2022 is co-held with EvaLatin2022. As an evaluation of the same type, EvaHan2022 has its own features. EvaLatin2022 mainly evaluates the NLP tools for Latin about Lemmatization and Part-of-Speech tagging. These 2 tasks are each with 3 sub-tasks (i.e. Classical, Cross-Genre and Cross-Time). Articles by five representative Latin authors were selected as Training data and Test data. Each team conducts a closed modality and then chooses whether to conduct an open modality. A total of five teams submitted test results in EvaLatin2022 (Sprugnoli et al., 2022), choosing the different methods and all the results exceed the baseline.

The best results in the lemmatization task for the three subtasks in terms of F1 score are 97.26% (Classical), 96.03% (Cross-genre) and 92.15% (Cross-time). And the best results in the POS tagging task for the three subtasks in terms of F1 score are 97.99% (Classical), 96.78% (Cross-genre) and 92.97% (Cross-time), as shown in Table 16. Also, we can see that the best results are almost all in open modality. Differently, EvaHan2022 divides the results of evaluation into four categories as *TestA* Closed modality, *TestA* Open modality, *TestB* Closed modality and *TestB* Open modality. The best results for these four types of tasks are 92.05% (FDU), 92.56% (FDU), 87.77% (FDU) and 89.47% (ZNNU).

Test	LF	PF
Classical Closed	96.45	97.61
Classical Open	97.26	97.99
Cross-Genre Closed	93.05	94.78
Cross-Genre Open	96.03	96.78
Cross-Time Closed	91.68	92.97
Cross-Time Open	92.15	92.70

Table 16. The best F1 scores on **L**emmatization(LF) and **P**OS tagging(PF) in EvaLatin2022 (%)

The shared tasks of EvaLatin2022 and EvaHan2022 both achieved good results. The POS tagging results of Latin are 4-5 points higher than that of Ancient Chinese. From the linguistic perspective, the inflections are the markers of the words' grammatical functions, thus the POS tagging of Latin is easier than Ancient Chinese. On the other hand, the best score of lemmatization of Latin is similar to that of word segmentation of Ancient Chinese, which is around 96%.

Comparing with the Mandarin Chinese's word segmentation and POS tagging scores in SIGHAN bakeoffs, the Ancient Chinese is around 1 point lower in word segmentation, while about 3 points lower in POS tagging.

6. Conclusion

EVAHan2022 is the first bakeoff for Ancient Chinese word segmentation and POS tagging. The best system from Fudan University outperforms almost all the other systems. Deep learning models raise up the scores for the Ancient Chinese, as it does on other languages like Latin. However, performance on single-source (ie. one book) dataset is better than on multiple-source datasets. It is caused by out-of-vocabulary (OOV) words in the new dataset. OOV is always a challenge for any lexical analyzers. So, there should be more attention paid to it.

In the future, the next EvaHan bakeoff should be extended to more genres and cross-time corpora, in order to improve the performance on more data.

7. Acknowledgements

Thank the reviewers for their advices. Thank Tongzheheng Zheng, Xue Yu, Min Shi, Lili Yu, Qingqing Wang, Yaxin Li, Qian Yang, Lu Wang, for their data annotation of the datasets. This work is supported in part by National Social Science Funds of China (18BYY127, 21&ZD331), Project of Social Science Foundation of Jiangsu Province (20JYB004).

8. Bibliographical References

Hongmei Zhao and Qun Liu. (2010). The CIPS-SIGHAN CLP2010 Chinese Word Segmentation Backoff. In *Proceedings of The First CIPS-SIGHAN Joint Conference on Chinese Language Processing.* pp. 199-209.

Huiming Duan., Zhifang Sui., Ye Tian and Wenjie Li. (2012). The CIPS-SIGHAN CLP 2012 ChineseWord Segmentation on MicroBlog Corpora Bakeoff. In *CIPS-SIGHAN*. pp.35-40.

Guangjin Jin and Xiao Chen. (2008). The Fourth International Chinese Language Processing Bakeoff: Chinese Word Segmentation, Named Entity Recognition and Chinese POS Tagging. In *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing*.

Gina-Anne Levow. (2006). The Third International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition. In *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*, pp.108–117, Sydney, Australia. Association for Computational Linguistics.

Ning Cheng, Bin Li, Liming Xiao, Changwei Xu, Sijia Ge, Xingyue Hao, Minxuan Feng. Integration of Automatic Sentence Segmentation and Lexical Analysis of Ancient Chinese based on BiLSTM-CRF Model. In *First Workshop on Language Technologies for Historical and Ancient Languages, (LT4HALA 2020)*, pp 52-58. Marseille, 11–16 May 2020.

Nianwen Xue. (2003). Chinese Word Segmentation as Character Tagging. *International Journal of Computational Linguistics & Chinese Language Processing,* Volume 8, Number 1, February 2003: Special Issue on Word Formation and Chinese Language Processing, pages 29–48.

Richard Sproat and Thomas Emerson. (2003). The first international Chinese word segmentation Bakeoff. In *Proceedings of the second SIGHAN workshop on Chinese language processing - Volume 17* (*SIGHAN '03*). Association for Computational Linguistics, USA, 133–143.

Sprugnoli, Rachele and Passarotti, Marco and Cecchini, Flavio Massimiliano and Fantoli, Margherita and Moretti, Giovanni. (2022). Overview of the EvaLatin 2022 evaluation campaign. In *Proceedings of LT4HALA*

2022-2st Workshop on Language Technologies for Historical and Ancient Languages.

Maosong Sun, Jiayan Zou.(2001). A Critical Appraisal of the research on Chinese Word Segmentation. China : *Contemporary Linguistics*, 2001, 3(1), 22-32+77.

Min Shi, Bin Li and Xiaohe Chen. (2010). CRF Based Research on a Unified Approach to Word Segmentation and POS Tagging for Pre-Qin Chinese. China: *Journal of Chinese Information Processing*. pp. 39-46.

Thomas Emerson. (2005). The Second International Chinese Word Segmentation Bakeoff. In *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing*.

Xiaohe Chen, Minxuan Feng, Runhua Xu. (2013). *Information Processing of Pre-Qin Chinese*. Beijing : World Publishing Corporation.

9. Language Resource References

Xiaohe Chen, Bin Li, Minxuan Feng, et al. (2017). Ancient Chinese Corpus, distributed via LDC, Text resources, 1.0, ISLRN 924-985-704-453-5.