Enhancing Ancient Chinese Understanding with Derived Noisy Syntax Trees

Shitou Zhang[1,2], **Ping Wang**[1,*], **Zuchao Li**[2,*], **Jingrui Hou**[3]
[1]School of Information Management, Wuhan University
[2]School of Computer Science, Wuhan University
[3]Department of Computer Science, Loughborough University
{shitouzhang,wangping,zcli-charlie}@whu.edu.cn
J.Hou@lboro.ac.uk

Abstract

Despite the rapid development of neural-based models, syntax still plays a crucial role in modern natural language processing. However, few studies have incorporated syntactic information into ancient Chinese understanding tasks due to the lack of syntactic annotation. This paper explores the role of syntax in ancient Chinese understanding based on the noisy syntax trees from unsupervised derivation and modern Chinese syntax parsers. On top of that, we propose a novel syntax encoding component – confidence-based syntax encoding network (cSEN) to alleviate the side effects from the existing noise caused by unsupervised syntax derivation and the incompatibility between ancient and modern Chinese. Experiments on two typical ancient Chinese understanding tasks, ancient poetry theme classification and ancient-modern Chinese translation, demonstrate that syntactic information can effectively enhance the understanding of ancient Chinese over strong baselines, and that the proposed cSEN plays an important role in noisy scenarios.

1 Introduction

Ancient Chinese literature, such as classical poetry, books, and records, is a highly representative and distinctive cultural heritage that is receiving increasing attention from the NLP academia. However, directly applying modern Chinese processing methods to ancient texts is not appropriate due to the differences in syntax and semantics between ancient and modern Chinese. Chinese is one of the oldest written languages in the world, with a history of at least 6,000 years (Norman, 1988). Over time, the language has undergone many changes, such as the transition from literary to vernacular Chinese in the early 20th century (Weiping, 2017), resulting in a significant gap between ancient and modern Chinese.

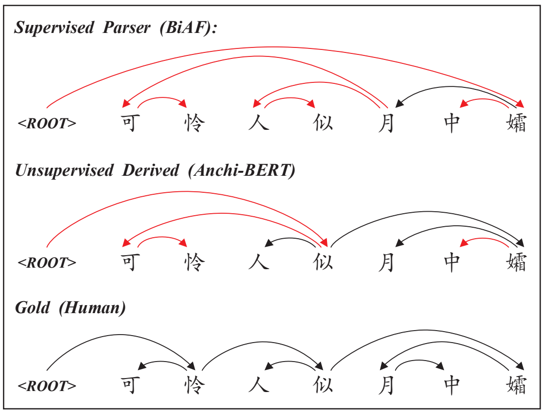

Figure 1: Unlabeled dependency parses from different parsers, where red arcs indicate prediction noises.

Syntactic features has been utilized in a wide range of NLP tasks, including coreference resolution (Fang and Fu, 2019; Trieu et al., 2019; Jiang and Cohn, 2022), machine reading comprehension (Zhang et al., 2020; Guo et al., 2020), and machine translation (Currey and Heafield, 2019; Zhang et al., 2019a; Bugliarello and Okazaki, 2020). Despite the effectiveness of syntax in modern Chinese understanding (Li et al., 2018; Xia et al., 2019; Zhang et al., 2020), few studies have incorporated syntactic information into ancient Chinese processing. Most works only take into account explicit features, such as era (Chang et al., 2021) and imagery (Shen et al., 2019), ignoring implicit syntactic features. The main reason for this lies in two aspects: (1) the linguistic gap between ancient and modern Chinese makes it difficult for supervised modern Chinese syntax parsers to correctly parse ancient Chinese expressions; (2) training a supervised ancient Chinese syntax parser from scratch can be highly costly due to the lack of annotated data.

Unsupervised syntax parsing or directly employing modern Chinese parsers will inevitably cause noise and performance degradation. A unlabeled example and corresponding human annotation on ancient Chinese sentence "可怜人似月中孀(*It is*

* Corresponding authors.

Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics - Student Research Workshop, pages 83–92
July 10-12, 2023

pitiful like Chang'e in the moon)" are shown in Figure 1. To address this challenge, we propose a novel syntax encoding structure – confidence-based syntax encoding network (cSEN), which alleviates the negative effect of noise by measuring confidence of arcs in syntax graphs. Specifically, confidence is calculated by performing Biaffine transformation over the sequence representation and the derived syntactic graph adjacency matrix. With this obtained confidence, our model is capable of distinguishing useful syntactic features from noise.

Moreover, compared with modern Chinese, ancient Chinese has more concise expressions and thus more compact structures, each token is highly relative to the preceding and following one. Considering such linguistic characteristic, we incorporate another graph feature – left-right branch (LRB), which captures local features to further improve ancient Chinese understanding. Experiments are conducted on two typical ancient Chinese understanding tasks, thematic classification of ancient poetry and ancient-modern Chinese translation. Results show that our model achieves significant improvements over powerful baselines, and our proposed cSEN can effectively handle the noise in the derived syntax trees. To our best knowledge, our proposed cSEN is the first solution that makes the syntax practical in ancient Chinese processing. The proposed cSEB can serve as a backbone for enriching our understanding of ancient texts, offering a scalable and consistent solution for education, research, and broadening the public's access to these significant cultural treasures.

Overall, the contributions of this paper can be concluded in four folds:

- This study fills the research gap of exploring the role of syntax in ancient Chinese understanding. Our work demonstrates that syntactic information, even noisy parses from unsupervised derivation, can benefit ancient Chinese understanding substantially.

- We propose a novel architecture – confidence-based syntax encoding network (cSEN), which alleviates the negative effect of noise in syntax parses, thus making it practical to utilize derived syntactic information to enhance ancient Chinese understanding.

- The effectiveness of cSEN is evaluated on two typical ancient Chinese understanding tasks, ancient poetry thematic classification and ancient-modern Chinese translation. Results show that our model yields significantly better performance in noisy scenarios over powerful baselines.

- We create a new dataset for the thematic classification of ancient Chinese poetry, with 22,360 poems divided into 10 theme categories. This dataset offers a data foundation for related research and helps to eliminate the lack of available ancient Chinese annotated corpora.

2 Related Work

2.1 Syntax Role in Modern Chinese Understanding

As syntax is highly correlated with semantics, syntactic features, including constituent and dependency structures, have been utilized in many modern Chinese understanding tasks and have been shown to be helpful clues. Li et al. (2018) explored the effect of syntax on semantic role labeling (SRL) and confirmed that high-quality syntactic parsing can effectively enhance syntactically-driven SRL. Xia et al. (2019) designed a syntax-aware multi-task learning framework for Chinese SRL by extracting implicit syntactic representations as external inputs for the SRL model. Jiang et al. (2018) incorporated syntactic features to expand identified triplets for improving Chinese entity relation extraction. Zhang et al. (2020) proposed a syntax-aware approach for solving machine reading comprehension, which incorporates explicit syntactic constraints into the attention mechanism for better linguistically motivated word representations. Sun et al. (2022) utilized syntactic features, which capture depth-level structure information, including non-consecutive words and their relations, to enhance recognition of Chinese implicit inter-sentence relations. Zhu et al. (2022) incorporated syntactic dependency information to determine entity boundaries for improving Chinese named entity recognition. Despite the increasing attention that syntax is receiving in modern Chinese understanding, few studies have attempted to utilize syntactic features for ancient Chinese understanding.

2.2 Ancient-Modern Chinese Translation

Unlike bilingual translation tasks, such as Chinese-English, ancient and modern Chinese are written using the same characters. Despite that, translating between ancient and modern Chinese can still be challenging for native speakers. This is due to two factors: (1) the syntactical structure and grammatical order of ancient Chinese are different from those of modern Chinese, making ancient Chinese expressions more concise yet also more confusing; (2) ancient Chinese frequently employs allusion, metaphor, and symbolic imagery to implicitly evoke sensory and emotional experiences, which increases the complexity of disambiguating the intended message.

In recent years, advancements in deep learning have led to significant progress in neural machine translation. For example, Zhang et al. (2019b) proposed an unsupervised algorithm that constructs sentence-aligned ancient-modern pairs, and an end-to-end neural model with copying mechanism and local attention to translate between ancient and modern Chinese. Liu et al. (2019) applied RNN-based (Bahdanau et al., 2014) and Transformer-based (Vaswani et al., 2017) machine translation models to this task. Considering the monolingual nature of this task, Yang et al. (2021) utilized pre-trained model UNILM (Dong et al., 2019) and an ancient Chinese pre-trained model Guwen-BERT to enhance performance. Over time, the Chinese language has evolved a lot, resulting in different characteristics of ancient Chinese in different eras. To address this, Chang et al. (2021) proposed a time-aware translation method, where the model predicts both the translation results and its particular era, and uses the predicted chronological feature as auxiliary information to bridge the linguistic gap between Chinese language in different eras.

2.3 Classification of Ancient Chinese Poetry

Classification of ancient Chinese poetry provides a basis for higher-level tasks, such as sentiment or style controllable poetry generation (Yang et al., 2018; Chen et al., 2019; Shao et al., 2021). In the past, statistical features and machine learning algorithms were commonly used. For example, Hou and Frank (2015) proposed a weakly supervised sentiment classification approach, which created a sentiment lexicon based on Weighted Personalized PageRank (WPPR). Shen et al. (2019) incorporated imagery features for analyzing the sentiment of Tang Poetry. In recent years, neural classifiers have been introduced to the task and made remarkable progress in performance. For instance, Xuan et al. (2018) designed a poetry style recognition model by stacking a genetic algorithm with CNN, and Tang et al. (2020) combined CNN with a gated GRU for solving poetry sentiment classification.

3 Model

In this section, we describe architecture of the proposed cSEN. We first present a basic GAT encoder, then introduce our cSEN. The overview of cSEN is shown in Figure 2.

3.1 Vanilla GAT

GAT is often applied over a sentence encoder to extract graph-based representations of the input text. Given input token sequence $\mathcal{T} = \{t_1, t_2, \ldots, t_l\}$, l denotes the sequence length. The output of the sentence encoder is denoted as matrix $\mathcal{H} \in \mathbb{R}^{l \times n}$, where each row $h_i \in \mathbb{R}^n$ is the representation of token t_i.

With dependency structure of the input sequence from a syntax parser, we construct a dependency graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of tokens and $\mathcal{E}$ is the set of arcs. In the graph encoding, we employ the form of adjacency matrix to describe the graph, in which the positions with arcs and diagonal are assigned to ones, denoted as $\mathcal{M}^{(\text{dep})}$. Linear transformation is performed by multiplying the sentence representation $\mathcal{H}$ with a matrix $\mathcal{W} \in \mathbb{R}^{n \times n'}$ for feature extraction, where n' denotes the transformed feature dimension:

$$\mathcal{Z} = \mathcal{H}\mathcal{W}.$$

Then, a pair-wise attention operation is performed. For every pair $t_i, t_j \in \mathcal{V}$, it concatenates corresponding representations z_i and z_j, then takes the dot product with vector $a \in \mathbb{R}^{2n'}$ and applies a **LeakyReLU** activation function:

$$\mathcal{S}^{(\text{raw})}[i,j] = \textbf{LeakyReLU}([z_i \oplus z_j]^T a),$$

where $\oplus$ represents the concatenation operation, and $\mathcal{S}^{(\text{raw})}$ is a score matrix with the size of $(l \times l)$ that captures inter-node relations. To integrate the graph structure, the adjacency matrix $\mathcal{M}^{(\text{dep})}$ is used to constrain the function scope before a regular **Softmax** operation is performed. By doing this, each token can only attend to its head tokens and itself. The obtained attention weights matrix then is used for scaling the transformed sentence representation $\mathcal{Z}$ and calculating the final attentional output:

$$\mathcal{W}^{(\text{attn})} = \textbf{Softmax}(\mathcal{S}^{(\text{raw})} \times \mathcal{M}^{(\text{dep})}).$$

$$\mathcal{H}^{(\text{attn})} = \mathcal{W}^{(\text{attn})}\mathcal{Z}.$$

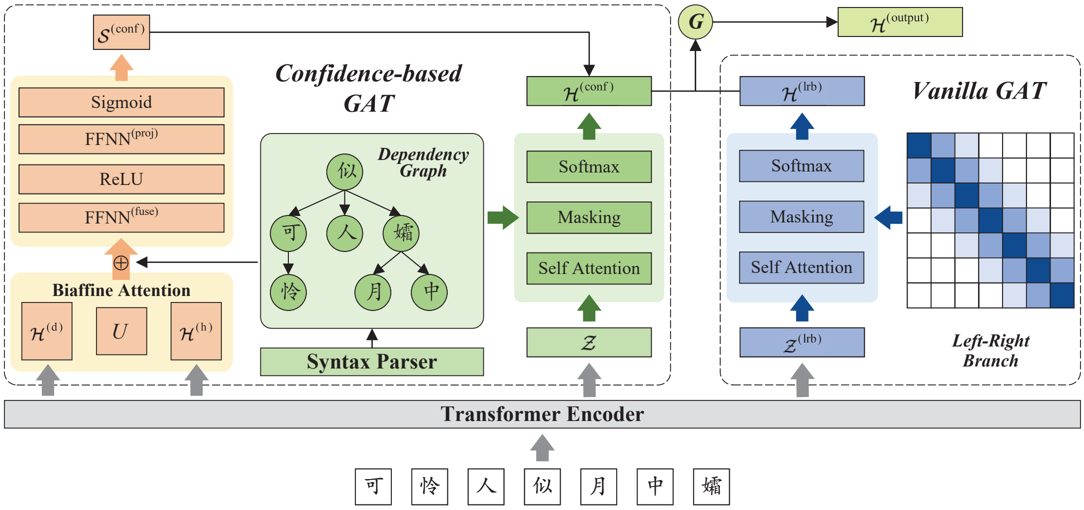

Figure 2: Architecture of the proposed cSEN. $\oplus$ and Ⓖ represents the concatenation operation and the gated mechanism, respectively. We present $\mathcal{M}^{(\text{dep})}$ in the form of graph where arcs are pointing from heads to dependencies. The cells in $\mathcal{M}^{(\text{lrb})}$ are colored to highlight the local dependencies, and darker color indicates higher correlation.

3.2 Confidence-based GAT

As discussed above, GAT guides the encoding process by constraining the scope of the attention computation. Therefore, the presence of noise in the graph will inevitably impact the encoding output. To alleviate the negative effects of noise on the model's performance, we propose a confidence-based GAT, which measures the confidence of the graph adjacency matrix, helping the model distinguish reliable syntactic information from noise.

Similar to vanilla GAT, we first model the pairwise relationships. Two separate linear transformations are performed over the sentence representation $\mathcal{H}$ to obtain the role-aware representations. The outputs are denoted as $\mathcal{H}^{(\text{d})}$ and $\mathcal{H}^{(\text{h})}$ respectively, both of which have the size of $(l \times n')$:

$$\mathcal{H}^{(\text{d})} = \mathcal{H}\mathcal{W}^{(\text{d})}; \mathcal{H}^{(\text{h})} = \mathcal{H}\mathcal{W}^{(\text{h})}.$$

Then, Biaffine attention (Dozat and Manning, 2016) are calculated on the role-aware representations for pair-wise relationship scoring:

$$\mathcal{S}^{(\text{bi})} = \mathcal{H}^{(\text{d})} U \mathcal{H}^{(\text{h})T},$$

where U is an intermediate matrix with the size of $(n' \times n')$. Confidence scores are calculated by concatenating the pair-wise relationship scores and the adjacency matrix and passing them through processing as follows,

$$\mathcal{S}^{(\text{fuse})} = \textbf{ReLU}(\textbf{FFNN}^{(\text{fuse})}(\left[\mathcal{S}^{(\text{bi})} \oplus \mathcal{M}^{(\text{dep})}\right])),$$

$$\mathcal{S}^{(\text{conf})} = \textbf{Sigmoid}(\textbf{FFNN}^{(\text{proj})}(\mathcal{S}^{(\text{fuse})})).$$

where $\textbf{FFNN}^{(\text{fuse})}$ performs a linear transformation to fuse the two feature spaces along with an **ReLU** activation, and $\textbf{FFNN}^{(\text{proj})}$ is used to reduce the dimension from $2l$ to l, so that **Sigmoid** can be applied to project the confidence features to the same magnitude as the attention scores. With this obtained confidence scores $\mathcal{S}^{(\text{conf})}$, we can remedy the original attention restrain process:

$$\mathcal{W}^{(\text{conf})} = \textbf{Softmax}(\mathcal{W}^{(\text{attn})} + \mathcal{S}^{(\text{conf})}),$$

$$\mathcal{H}^{(\text{conf})} = \mathcal{W}^{(\text{conf})} \mathcal{Z}.$$

In summary, cSEN alleviates the negative effect of noise in graphs through a two-fold process. First, cSEN measures the confidence of the derived syntax parses. This confidence score is then used to soft-mask noisy arcs and highlight previously undetected ones. Second, considering the linguistic characteristics of ancient Chinese, the Left-Right Branch feature is incorporated to broaden the scope of syntax graph encoding and smooth out noise and incompatibility. The combined effect of these aspects helps alleviate performance degradation caused by noise.

3.3 Left-Right Branch Feature

Inspired by the ubiquity of local dependencies in ancient Chinese, we introduce a novel straightforward and effective feature, left-right branch, to further improve the GAT encoding. To model local inter-token relations, we populate a matrix $\mathcal{M}^{(\text{lrb})}$ of the same size as $\mathcal{M}^{(\text{dep})}$ following

$$\mathcal{M}^{(\text{lrb})}[i, j] = \begin{cases} 1, & \text{if } j \in \{i-1, i+1\} \\ 0, & \text{otherwise.} \end{cases}$$

This indicates that there exist arcs in the graph connecting the node and its close left and right neighbors. The left-right branch features are encoded using another GAT component, yielding a sequence representation $\mathcal{Z}^{(lrb)}$ and a positional-information-introduced attention weight matrix $\mathcal{W}^{(\mathrm{lrb})}$. The outputs from $\mathcal{M}^{(dep)}$ and $\mathcal{M}^{(lrb)}$ are combined with a gated mechanism to produce the final output:

$$\mathcal{H}^{(\mathrm{lrb})} = \mathcal{W}^{(\mathrm{lrb})}\mathcal{Z}^{(\mathrm{lrb})}.$$

$$g = \textbf{Sigmoid}(\textbf{FFNN}^{(\mathrm{gate})}(\left[\mathcal{H}^{(\mathrm{conf})} \oplus \mathcal{H}^{(\mathrm{lrb})}\right])),$$

$$\mathcal{H}^{(\mathrm{output})} = g \times \mathcal{H}^{(\mathrm{conf})} + (1-g) \times \mathcal{H}^{(\mathrm{lrb})}.$$

4 Experiments

We evaluate the effectiveness of cSEN module using two typical ancient Chinese understanding tasks: Thematic classification of ancient poetry and ancient-modern Chinese translation. We build our model by incorporating the cSEN module into existing solid baselines. For the classification task, we follow the work of (Vaibhav et al., 2019) which has a BERT-GAT-BiLSTM backbone architecture. And for the translation task, our model is based on (Jin et al., 2020) where dependency graphs are incorporated into neural sequence-to-sequence models with a pointer network.

4.1 Data

To address the scarcity of annotated data for thematic classification, we constructed a novel dataset[1]. Two graduate students specializing in Chinese literature study annotated 22,360 poems, categorizing them into one of ten distinct themes under the guidance of an experienced ancient Chinese linguist. This meticulous process ensured high-quality, reliable annotations. Any conflicted labelling between the two annotators was resolved through consultation with the supervisor, guaranteeing a consistent annotation standard. The dataset is then randomly divided into a training set (20,360), a development set (800), and a test set (1,200). The distribution of themes in the dataset is detailed in Table 1.

	Train	Dev	Test
#Object-chanting	1129	47	66
#Landscape	1097	44	47
#Persons	2403	91	129
#History	1087	40	76
#Homesickness	9013	357	522
#Mourning	503	18	31
#War	1746	62	115
#Pastoral	1219	47	84
#Farewell	1460	60	83
#Boudoir-plaint	703	34	47
Total	20360	800	1200

Table 1: Data statistics of the ancient Chinese poetry thematic classification dataset

For the ancient-modern Chinese translation, we adopt the ancient-modern Chinese parallel corpus contributed by the open source NiuTrans project[2]. The corpus contains 967,255 sentence pairs extracted from ancient Chinese books. We divided the corpus into training, validation, and test sets with corresponding sizes of 900,000, 60,000, and 7,255.

4.2 Syntax Parsing

We experiment with two settings – modern supervised parsers and ancient unsupervised syntax derivation. For modern supervised parsing, we adopt the Biaffine dependency parse (Dozat and Manning, 2016) and train it on CTB7 (Xue et al., 2010). For unsupervised syntax derivation, we follow the work of Wu et al. (2020), which utilizes linguistic knowledge gained from pre-trained language model BERT to infer syntactic dependency structure without direct supervison. We attempt two variants of BERT for syntax derivation and backbone sentence encoder, BERT-wwm-ext (Cui et al., 2021) and Anchi-BERT (Tian et al., 2021). BERT-wwm-ext is trained on the modern Chinese corpus containing 5.4B words, while Anchi-BERT is trained upon a ancient Chinese corpus with the size of 39.5M tokens. In addition, we treat the left-right branch as a special kind of syntax parses. Anchi-BERT is trained on a smaller ancient Chinese corpus (39.5M tokens), while BERT-wwm-ext is trained on a larger modern Chinese corpus (5.4B tokens). We also treat left-right branch features as a distinct class of syntax parses.

For clarity, the syntactic parses from the Biaffine parser, BERT-wwm derivation, and Anchi-BERT derivation are denoted as *BiAF*, *WWMD*, *ANCD* respectively, in the following part.

4.3 Implementation and Hyper-parameters

For the thematic classification, our model is built by stacking BERT, a graph encoder, and a single-layer LSTM. For the baseline, we do not incorporate syn-

[1] Upon publication of this paper, this dataset will be made available for research purposes.

[2] https://github.com/NiuTrans/Classical-Modern

Methods	**Parses**	BERT-wwm Micro F1	BERT-wwm Macro F1	Anchi-BERT Micro F1	Anchi-BERT Macro F1
Baseline	None	91.7	89.2	92.4	90.4
GAT	*LRB*	91.5	88.9	93.3	91.4
	BiAF	92.3	89.7	93.3	91.2
	WWMD	91.4	88.8	92.7	90.8
	ANCD	91.8	89.2	93.2	91.0
	BiAF+LRB	92.7	90.4	93.3	91.2
	WWMD+LRB	91.7	89.6	93.2	91.2
	ANCD+LRB	90.8	88.2	92.8	90.7
	BiAF+ANCD+LRB	91.7	88.8	92.6	90.6
cSEN	*BiAF+LRB*	91.4	89.2	93.3	91.6
	WWMD+LRB	92.8	90.7	93.6	**91.9**
	ANCD+LRB	91.3	89.1	93.2	91.3
	BiAF+ANCD+LRB	91.0	89.1	**93.8**	**91.9**

Table 2: Comparison with baseline model and syntax-aware methods on the thematic classification task.

Methods	**Parses**	BLEU	RG-1 F-score	RG-2 F-score	RG-L F-score
Baseline	None	37.14	69.71	46.24	67.62
GAT	*LRB*	37.42	69.86	46.36	67.72
	BiAF	37.45	70.23	46.93	68.21
	WWMD	37.46	70.20	46.89	68.14
	ANCD	37.55	69.90	46.53	67.85
	BiAF+ANCD+LRB	34.62	69.20	45.15	67.15
cSEN	*BiAF+ANDC+LRB*	**37.73**	**70.27**	**47.09**	**68.23**

Table 3: Experimental Results of the ancient and modern Chinese translation task.

tax parses, rendering the graph encoder ineffective in shaping the attention scope. The graph encoder's node embedding dimension is set to 128, and the hidden size in LSTM is set to 100. We adopt the Adam optimizer with $\rho = 5e-5$ and $\epsilon = 1e-8$, using a batch size of 32. All classifiers are trained for 10 epochs on the train set by default.

We mostly follow the parameter settings from (Jin et al., 2020) for the ancient-modern Chinese translation. The Adam optimizer is configured with $\rho = 1e-4$ and $\epsilon = 1e-8$. And all models are trained for 50 epochs with a batch size of 108.

4.4 Results

4.4.1 Ancient Poetry Thematic Classification

Table 2 presents the results of ancient poetry thematic classification. We report the results in Micro-F1 and Macro-F1 scores. The table is divided into three blocks, showing the results of the baseline model, vanilla GAT, and the proposed cSEN. The baseline model achieves 92.4 in Micro F1 and 90.4 in Macro F1, showing strong performance.

From the results in the first two blocks, it can be found that incorporating syntactic trees with GAT encoder brings substantial improvement, proving the value of syntactic information for enhancing ancient Chinese understanding. Through comparing the results of employing Anchi-Bert as the sentence encoder and those obtained employing Bert-wwm, we can see that Anchi-Bert outperforms BERT-wwm with a significant lead in all cases. Recall that Anchi-Bert was pre-trained on a much smaller corpus. Also, the performance of syntactic trees derived by BERT-wwm is inferior to the other three. This once more indicates the linguistic gap and syntactic incompatibility between ancient and modern Chinese.

Unsupervised syntax trees derived by Anchi-BERT performs roughly the same as those produced by the Biaffine parser. Additionally, LRB is the best-performing syntax parse among all, improving the performance by 0.9 in Micro F1 and 1.0 in Macro F1. It can be partially explained by the fact that ancient poems are comprised by a few brief sentences, which are highly concise and structurally compact. This results in fewer long-range dependencies, and each token is closely dependent on the immediate preceding or succeeding token.

From the third block, it can be seen that when using Anchi-BERT as sentence encoder, cSEN brings

Variants	Micro F1	Macro F1
cSEN	**93.8**	**91.9**
w/o Confidence	92.8	91.1
w/o Gate	93.0	91.0

Table 4: Ablation study results.

Syntax Trees	Micro F1	Macro F1
[*ANCD*] + (*LRB*)	93.2	91.3
[*BiAF*] + (*LRB*)	93.3	91.6
[*BiAF* + *LRB*] + (*ANCD*)	92.8	90.9
[*ANCD* + *LRB*] + (*BiAF*)	92.8	90.5
[*BiAF* + *ANCD*] + (*LRB*)	**93.8**	**91.9**

Table 5: Comparison of different combination configurations on syntactic parses. Parses in square brackets are merged onto a single adjacency matrix and parses in parentheses are incorporated by the gated mechanism

performance gains across all syntax trees setups, raising the top Micro and Macro F1 scores to 93.8 and 91.9, respectively. This demonstrates that: (1) cSEN's denoising capability is effective for utilizing noisy syntactic information to improve ancient Chinese understanding; (2) cSEN can handle noise introduced by different parses, whether it is from a supervised modern Chinese parser or unsupervised derivation.

4.4.2 Ancient-Modern Chinese Translation

Results of the ancient-modern Chinese translation are shown in Table 3. We use BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004) scores for performance evaluation. The baseline model without syntax parses achieves 37.14 in BLEU score and F-scores of 69.71, 46.24, 67.62 in ROUGE-1, ROUGE-2, and ROUGE-L respectively. With single syntactic parses incorporated, all models achieve better performance in all metrics, proving that syntax can effectively improve ancient-modern Chinese translation. LRB is relatively the weakest one, slightly increasing BLEU score by 0.28, and ROUGE f-scores by 0.15, 0.12, 0.10. This might be caused by that sentences from the ancient books have more long-distance dependencies and more complicated syntactic structures that left right branch can not recover. Anchi-BERT derived syntax parses have better performance with an improvement of 0.41 in BLEU score, and 0.19, 0.29, and 0.23 in ROUGE scores. BERT-wwm derived syntax trees and trees generated by Biaffine parser have similar results. In contrast to Anchi-BERT derived trees, their performance are inferior in BLEU scores but better in ROUGE F-scores. Feeding multiple syntactic parses into the GAT-based model simultaneously leads to a significant performance drop. While replacing GAT with the proposed cSEN increases performance in all metrices, with 37.73 in BLEU score and 70.27, 47.09, 68.23 in ROUGE F-scores. From the above results, we conclude that syntax parses from unsupervised derivation or modern Chinese syntax parsers introduce noise and degrade model performance. With our confidence learning, model is able to distinguish and separate informative syntactic information from noise, thus alleviating its negative effect.

Table 6 shows three ancient-to-modern Chinese translation examples produced by different models. From generations for Sent 1, we can see a common error: due to the lack of contextual information, all three models assume the surname of "the father" useing the most common Chinese surnames, such as "Li" and "Zhang". For Sent 2, the generations from the baseline model and vanilla GAT differ significantly from the human-annotated reference. They fail to recognize the relationship between the characters,such as who "其娣" refers to, thus generating tranlations that did not correspond to the facts. In contrast, with stronger denoising capability, cSEN is able to correctly encodes the information in ancient Chinese texts, thus producing higher-quality translations.

5 Exploration

In this section, we investigate the impact of different cSEN components and analyze the nature of different syntax parses.

First, we conduct ablation studies on cSEN. Results are reported in Table 4. Both the removal of the confidence (equivalent to vanilla GAT) and the removal of the gated mechanism lead to significant performance degradation. This indicates that both modules are crucial for effectively utilizing syntactic information from noisy parses.

Second, we experiment with different syntax tree combination setups. Table 5 lists the findings. Combining two syntax parsers onto a single dependency graph can provide richer syntactic information and produce higher performance than using alone syntax parses, no matter it is from unsupervised derivation or a supervised parser. We also explore the incorporation position of LRB features. It can be observed that the model performance suffers significantly if LRB features and graph parses

Sent	Ancient Chinese
1	src: 父岘，少骁勇，事章大宝为帐内部曲。 ref: 父亲陈岘，年轻时骁勇，侍奉章大宝，任帐内部曲。 trl: His father, Chen Xian, was brave when he was young. He served Zhang Dabao, as an in-tent soldier.
2	src: 晋伐骊戎，骊戎男女以骊姬。归生奚齐。其娣生卓子。 ref: 晋国攻打骊戎，骊戎男把骊姬献给晋献公，回国后生了奚齐，她的妹妹生了卓子。 trl: Jin state attacked Li Rong. Li Rong’s men presented Li Ji to Jin Xian Gong. After returning the state, Li Ji gave birth to Xi Qi, and her sister gave birth to Zhuo Zi.
3	src: 及竣事，又曰：杨存中之罢，朕不安寝者三夕。 ref: 等到事毕后，又说：杨存中被罢免，我有三天不得安寝。 trl: After the matter was over, he said: "The dismissal of Yang Cunzhong, made me unable to sleep for three days."
Baseline	
1	父亲张岘，少年时骁勇善战，在章大宝帐下当兵, 任章大宝的部曲。 His father, Zhang Xian, was brave and good at fighting when he was young. He served as a soldier under the command of Zhang Dabao, and was Zhang Dabao’s soldier.
2	晋国讨伐骊戎，骊戎男女献上骊姬，回国后生了奚齐，骊戎的妹妹生了卓子。 Jin state attacked Li Rong, and Li Rong’s men and women presented Li Ji, and after returning the state, they gave birth to Xi Qi, and Li Rong’s sister gave birth to Zhuo Zi.
3	等到完成任务，又说：杨存中被罢免，朕三夜不安寝。 After completing the task, he said: "The dismissal of Yang Cunzhong, made me unable to sleep for three nights."
Vanilla GAT	
1	父亲李岘，少年时骁勇善战，在章大宝手下做事。 His father, Li Xian, was brave and good at fighting when he was young, and worked under Zhang Dabao.
2	晋国讨伐骊戎，骊戎男娶了骊姬，嫁给了奚齐，奚齐的妹妹生了卓子。 Jin state attacked Li Rong. Li Rong’s men married Li Ji, married Xi Qi, and Xi Qi’s sister gave birth to Zhuo Zi.
3	等到完工，又说：杨存中被罢免的事，我感到不安宁 的有三个晚上。 After the work was finished, he said: I felt uneasy about Yang Cunzhong’s dismissal for three nights.
cSEN	
1	父亲陈岘，年轻时骁勇，事奉章大宝任帐内部曲。 trl: His father, Chen Xian, was brave when he was young. He served Zhang Dabao as an in-tent soldier.
2	晋国攻打骊戎，骊戎男把骊姬送给晋国，回国后生了奚齐，她的妹妹生了卓子。 Jin state attacked Li Rong. Li Rong’s men presented Li Ji to Jin State. After returning the state, Li Ji gave birth to Xi Qi, and her sister gave birth to Zhuo Zi.
3	等到事情完毕，又说：杨存中被罢免，我三天不安寝。 trl: After the matter was over, he said: "The dismissal of Yang Cunzhong, made me unable to sleep for three days."

Table 6: Ancient-to-modern Chinese translation examples generated by the baseline model, vanilla GAT, and cSEN. The first block shows the original ancient Chinese sentence (src), human-annotated modern Chinese reference (ref), and corresponding English translations (trl).

are directly merged together. This again indicates the necessity of our gated method for LRB feature integration.

Third, as illustrated in Figure Figure 3, we compare our model and baselines over different input lengths. cSEN performs better in relative longer sentences, according to the results. This supports the hypothesis that syntax helps guide longer sentence understanding as dependency reduces the distance. Because of the incompatibility between modern and ancient Chinese, unsupervised derivation is more effective than supervised parsing when compared to other syntax parsers. In most cases, cSEN yeilding better performance due to its stronger denoising capabilities.

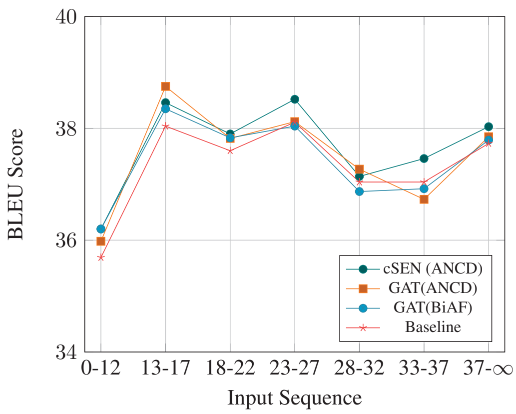

Figure 3: BLEU scores for different input sentence lengths.

6 Conclusions

In this paper, we investigate the role of syntax in improving ancient Chinese understanding. Due to lack of syntax annotation, syntax trees are obtained by unsupervised derivation and supervised modern Chinese parser. To alleviate the negative effect of noise, we propose a confidence-based syntax encoding network (cSEN). Experimental results on two typical ancient Chinese understanding tasks show that our model can effectively distinguish informative syntactic information from noise and achieve better performance. The application of our proposed cSEN can enhance the accessibility of ancient Chinese resources by offering a scalable and consistent solution for mining semantic information of ancient Chinese texts.

Limitations

The main limitation of our study comes from the extra parameters caused by confidence calculation, in which two separate self-attention operations and Biaffine transformation are performed. Incremental parameters results in a more time-consuming training process, and a higher hardware demand for storage. To address this issue, we plan to combine parameters from different attentional transformations into shared weight matrices in our future work to reduce the model size.

Acknowledgements

This paper was partially supported by the National Natural Science Foundation of China [No. 72074171].

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Emanuele Bugliarello and Naoaki Okazaki. 2020. Enhancing machine translation with dependency-aware self-attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1618–1627.

Ernie Chang, Yow-Ting Shiue, Hui-Syuan Yeh, and Vera Demberg. 2021. Time-aware ancient chinese text translation and inference. *arXiv preprint arXiv:2107.03179*.

Huimin Chen, Xiaoyuan Yi, Maosong Sun, Wenhao Li, Cheng Yang, and Zhipeng Guo. 2019. Sentiment-controllable chinese poetry generation. In *IJCAI*, pages 4925–4931.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Anna Currey and Kenneth Heafield. 2019. Incorporating source syntax into transformer-based neural machine translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 24–33.

Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. *Advances in Neural Information Processing Systems*, 32.

Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.

Kong Fang and J Fu. 2019. Incorporating structural information for better coreference resolution. In *Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19*.

Shaoru Guo, Yong Guan, Ru Li, Xiaoli Li, and Hongye Tan. 2020. Incorporating syntax and frame semantics in neural network for machine reading comprehension. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2635–2641.

Yufang Hou and Anette Frank. 2015. Analyzing sentiment in classical chinese poetry. In *Proceedings of the 9th SIGHUM workshop on language Technology for Cultural Heritage, social sciences, and humanities (LaTeCH)*, pages 15–24.

Fan Jiang and Trevor Cohn. 2022. Incorporating constituent syntax for coreference resolution. *arXiv preprint arXiv:2202.10710*.

Yishun Jiang, Gongqing Wu, Chenyang Bu, and Xuegang Hu. 2018. Chinese entity relation extraction based on syntactic features. In *2018 IEEE International Conference on Big Knowledge (ICBK)*, pages 99–105. IEEE.

Hanqi Jin, Tianming Wang, and Xiaojun Wan. 2020. Semsum: Semantic dependency guided neural abstractive summarization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8026–8033.

Zuchao Li, Shexia He, Jiaxun Cai, Zhuosheng Zhang, Hai Zhao, Gongshen Liu, Linlin Li, and Luo Si. 2018. A unified syntax-aware framework for semantic role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2401–2411.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Dayiheng Liu, Kexin Yang, Qian Qu, and Jiancheng Lv. 2019. Ancient–modern chinese translation with a new large training dataset. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(1):1–13.

Jerry Norman. 1988. *Chinese*. Cambridge University Press.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Yizhan Shao, Tong Shao, Minghao Wang, Peng Wang, and Jie Gao. 2021. A sentiment and style controllable approach for chinese poetry generation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 4784–4788.

Yabo Shen, Yong Ma, Chunguo Li, Shidang Li, Mingliang Gu, Chaojin Zhang, Yun Jin, and Yingli Shen. 2019. Sentiment analysis for tang poetry based on imagery aided and classifier fusion. In *International Conference on Artificial Intelligence for Communications and Networks*, pages 283–290. Springer.

Kaili Sun, Yuan Li, Huyin Zhang, Chi Guo, Linfei Yuan, and Quan Hu. 2022. Syntax–aware graph convolutional network for the recognition of chinese implicit inter-sentence relations. *The Journal of Supercomputing*, pages 1–24.

Yongrui Tang, Xumei Wang, Peng Qi, and Yan Sun. 2020. A neural network-based sentiment analysis scheme for tang poetry. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 1783–1788. IEEE.

Huishuang Tian, Kexin Yang, Dayiheng Liu, and Jiancheng Lv. 2021. Anchibert: a pre-trained model for ancient chinese language understanding and generation. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Hai Long Trieu, Anh-Khoa Duong Nguyen, Nhung Nguyen, Makoto Miwa, Hiroya Takamura, and Sophia Ananiadou. 2019. Coreference resolution in full text articles with bert and syntax-based mention filtering. In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 196–205.

Vaibhav Vaibhav, Raghuram Mandyam Annasamy, and Eduard Hovy. 2019. Do sentence interactions matter? leveraging sentence level representations for fake news classification. *arXiv preprint arXiv:1910.12203*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Chen Weiping. 2017. An analysis of anti-traditionalism in the new culture movement. *Social Sciences in China*, 38(2):175–187.

Zhiyong Wu, Yun Chen, Ben Kao, and Qun Liu. 2020. Perturbed masking: Parameter-free probing for analyzing and interpreting bert. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4166–4176.

Qingrong Xia, Zhenghua Li, and Min Zhang. 2019. A syntax-aware multi-task learning framework for chinese semantic role labeling. *arXiv preprint arXiv:1911.04641*.

Jing Xuan, Zhongshi He, Liangyan Li, Weidong He, Fei Guo, Hang Zhang, and Qiong Wu. 2018. Brain-oriented cconvolutional neural network computer style recognition of classical chinese poetry. *NeuroQuantology*, 16(4).

Nianwen Xue, Zixin Jiang, Xiuhong Zhong, Martha Palmer, Fei Xia, Fu-Dong Chiou, and Meiyu Chang. 2010. Chinese treebank 7.0. `https://catalog.ldc.upenn.edu/LDC2010T07`. Accessed: 2022-05-20.

Cheng Yang, Maosong Sun, Xiaoyuan Yi, and Wenhao Li. 2018. Stylistic chinese poetry generation via unsupervised style disentanglement. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3960–3969.

Zinong Yang, Ke-jia Chen, and Jingqiang Chen. 2021. Guwen-unilm: Machine translation between ancient and modern chinese based on pre-trained models. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 116–128. Springer.

Meishan Zhang, Zhenghua Li, Guohong Fu, and Min Zhang. 2019a. Syntax-enhanced neural machine translation with syntax-aware word representations. *arXiv preprint arXiv:1905.02878*.

Zhiyuan Zhang, Wei Li, and Qi Su. 2019b. Automatic translating between ancient chinese and contemporary chinese with limited aligned corpora. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 157–167. Springer.

Zhuosheng Zhang, Yuwei Wu, Junru Zhou, Sufeng Duan, Hai Zhao, and Rui Wang. 2020. Sg-net: Syntax-guided machine reading comprehension. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9636–9643.

Peng Zhu, Dawei Cheng, Fangzhou Yang, Yifeng Luo, Dingjiang Huang, Weining Qian, and Aoying Zhou. 2022. Improving chinese named entity recognition by large-scale syntactic dependency graph. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:979–991.