Approaches and Challenges for Resolving Different Representations of Fictional Characters for Chinese Novels

Li Song[1,2], **Ying Liu**[2]
[1] School of Humanities and Arts, Tianjin University, Tianjin 300072, China
[2] Department of Chinese Language and Literature, Tsinghua University, Beijing 100084, China
songli1105@sina.com, yingliu@tsinghua.edu.cn

Abstract

Due to the huge scale of literary works, automatic text analysis technologies are urgently needed for literary studies such as Digital Humanities. However, the domain-generality of existing NLP technologies limits their effectiveness on in-depth literary studies. It is valuable to explore how to adapt NLP technologies to the literary-specific tasks. Fictional characters are the most essential elements of a novel, and thus crucial to understanding the content of novels. The prerequisite of collecting a character's information is to resolve its different representations. It is a specific problem of anaphora resolution which is a classical and open-domain NLP task. We adapt a state-of-the-art anaphora resolution model to resolve character representations in Chinese novels by making some modifications, and train a widely used BERT fine-tuned model for speaker extraction as assistance. We also analyze the challenges and potential solutions for character-resolution in Chinese novels according to the resolution results on a specific Chinese novel.

Keywords: anaphora resolution, speaker extraction, fictional character, Chinese novel

1. Introduction

Due to the huge scale of literary works, relying only on human reading can hardly meet the understanding and application needs. Automatic text analysis technologies are urgently needed for literary studies such as Digital Humanities. However, existing automatic NLP tools and models are often developed for generic, open-domain tasks, restricting in-depth literary studies. Novel is probably the most important literary genre. Fictional characters are the most essential elements of a novel, and behaviors of characters promote the development of the plot. Therefore, characters play a crucial role in understanding the content of novels. However, representations of characters are complex and diverse in novels, causing much trouble for automatic analysis. In addition to proper nouns such as names, aliases and nicknames, e.g. 白嘉轩 (Bai Jia-xuan), characters are also frequently represented by pronouns, e.g. 他 (he), and anaphoric noun phrases, e.g. 头胎儿子 (the first-born son). Moreover, because Chinese is a pro-drop language, characters also often exist in the form of zero pronouns (ZPs), e.g. the "∅" in 田福贤起初愣了半刻，随之∅就打断了鹿子霖的话 (At first Tian Fu-xian was taken aback for a moment, then ∅ interrupted Lu Zi-lin) refers to Tian Fu-xian. The prerequisite of collecting a character's information is to resolve its different representations. It is a specific problem of anaphora resolution, which is a classical and open-domain NLP task. In this paper, we try to adapt a state-of-the-art anaphora resolution model to resolve character representations in Chinese novels by making some modifications, and train a BERT fine-tuned model for speaker extraction as assistance. Additionally, we test our approaches on a specific Chinese novel, and analyze the challenges and potential solutions for character-resolution in Chinese novels according to the resolution results.

Because most languages rarely allow pro drop[1], anaphora resolution in NLP is usually divided into two subtasks, zero pronoun resolution and non-zero coreference resolution. Zero pronoun resolution usually comprises two steps: recognizing dropped anaphoric pronouns, and then finding out their anaphoric mentions, namely occurrences of the same entity. Non-zero coreference resolution targets at recognizing mentions and clustering those referring to the same entity. Error propagation is inevitable if zero and non-zero anaphora is resolved step by step. Besides, it is another challenge to integrate zero and non-zero anaphora resolution results for the same text. Chen et al. (2021) proposed an end-to-end neural model[2] for tackling the two tasks jointly for Chinese (CHEN-Model for short), that is, it avoids decomposing the anaphora resolution task and integrating the resolution results. This model applied on the OntoNotes 5.0 Chinese dataset[3] (OntoNotes-*Ch* for short) can achieve state-of-the-art performance on both tasks. Considering the text characteristics and application requirements of Chinese novels, we make some modifications on CHEN-Model to design a character resolution model for Chinese novels: **A**naphora **R**esolution **M**odel for fictional **C**haracters in **C**hinese novels (ARMCC for short). Experimental results show that the F_1-score of ARMCC is 70.62% and 32.25% for the two subtasks. The training process and performance of ARMCC can be seen in Section 3.

Dialogs are useful for portraying characters and promoting the development of the plot, and dialogs between characters are an important part of novel texts. The word tokens in quotes make up 8% to 20% of all the word tokens in 240 Chinese novels of 12 genres (Chen et al., 2019). This percentage in Jin Yong's 15 novels is up to 40.24% (Jia et al., 2020). Additionally, our experiments show that CHEN-Model's performance on conversational texts is improved significantly when using speaker information, i.e. whether the two candidate

[1] https://wals.info/.
[2] https://github.com/cheniison/e2e-joint-coref.
[3] http://catalog.ldc.upenn.edu/LDC2013T19.

LREC-COLING 2024, pages 1408–1421

representations are from the same speaker. Therefore, we infer that speaker information is useful to resolve character representations in novels. However, novels are not one of the genres in OntoNotes-*Ch*, which has provided speaker information. We therefore train a SQuAD v1.1 BERT model [4] (Devlin et al., 2019) for automatically extracting speakers of quotes for Chinese novels: **S**peaker **E**xtractor for **C**hinese **N**ovels (SECN for short). Given a piece of novel text containing a quote, the model directly extracts a span from the text as the answer to the question 'who is the speaker of the quote?'. We integrate datasets of Chen et al. (2019) and Jia et al. (2020) and convert them into the form of Stanford Question Answering Dataset (SQuAD) (Rajpurkar et al., 2016), then finetune the pre-trained model. Experiments demonstrate good performance of SECN with 96.80% exact match accuracy. The reconstruction scheme of the dataset and error analysis of SECN are presented in Section 4.

We then test our approaches on a famous Chinese novel *White Deer Plain* (Chen, 1993) and make a detailed result analysis. The resolution results indicate that our approaches can satisfy the needs of coarse-grained researches and greatly reduce human effort in fine-grained researches on characters in Chinese novels. We also summarize the difficulties and challenges to character-resolution in Chinese novels and provide potential solutions to some problems. See Section 5 for details.

2. Related Works

Chinese Zero Pronoun Identification aims to identify the positions of the dropped anaphoric pronouns in Chinese. It is a prerequisite step before ZP resolution without gold positions of ZPs. Chang et al. (2017) develop a two-layer stacked bidirectional LSTM model with raw text and part-of-speech information to tackle identification of Chinese ZPs. Aloraini and Poesio (2020) introduce BERT to capture structural properties of languages and propose a BERT-based multilingual model for identifying ZPs. Iwata et al. (2021) formalize ZP identification as a question-answer-based span prediction task, namely given a predicate as a query and predict the omission with ZPs.

Chinese Zero Pronoun Resolution is mostly based on gold position of anaphoric ZPs or needs to identify ZPs beforehand, so that focuses only on finding anaphoric mentions for the given ZPs. Chen and Ng (2016) apply a feed-forward neural network where ZPs are encoded by its previous word and headword to resolve Chinese ZPs. Yin et al. (2017) apply LSTM models to encode ZPs and their candidate antecedents, and apply attention mechanism to filter candidate antecedents. Lin and Yang (2020) design a two-layer attention model to generate representations for ZPs and candidate antecedents and integrate constraint of similarities among correct antecedents into pairwise-margin loss. There are also researchers exploring end-to-end approaches of jointly handling identification and resolution of Chinese ZPs. Song et al. (2020) present a BERT-based multi-task model which treats ZP identification as a sequence labeling task, regarding whether each gap between words is a ZP and what type it is, and formulates ZP resolution as a reading comprehension task, treating each gap as a query and its anaphoric mentions as the answers.

Non-zero Coreference Resolution is a language-general task and aims to recognize mentions and cluster those referring to the same entity. Lee et al. (2017) consider all spans in a document as potential mentions and first jointly tackle mention detection and coreference prediction in an end-to-end method. Lee et al. (2018) improve this approach in various ways including word embedding, span representation and score function. Kantor and Globerson (2019) provide Entity Equalization mechanism to enable the model to use global entity-level information. Wu et al. (2020) treat coreference resolution as question answering task and use a plethora of existing question answering datasets as data augmentation. There are also some efforts on tackling ZP resolution and non-zero coreference resolution jointly for pro-drop languages. Iida and Poesio (2011) propose an Integer Linear Programming-based model which integrates the zero-anaphora resolver with a coreference resolver. Chen et al. (2021) propose an end-to-end neural model for tackling the two tasks jointly for Chinese, and we take it as the base model to develop our character resolution model in this paper.

Quote Attribution targets at identifying speakers for each quote in the document. Muzny et al. (2017) present a deterministic sieve-based system that models quote attribution as a two-step process: identifying the mention that corresponds to the speaker of a quote, and then resolving the mention to an entity. Cuesta-Lazaro et al. (2022) formulate quote attribution in literary texts as a dialogue state tracking task by using a BERT-based model to track the speaker for every single quote in conversations. Wang et al. (2022) proposes a Rule-BertAtten method for quote attribution in Chinese novels. They classify quotes according to the features of candidate speakers and then apply a rule-based method and a method combining BERT with attention mechanism to different categories of quotes respectively.

3. Character Resolution Model ARMCC

3.1 Task Description

ARMCC aims to resolve all the character representations in the form of ZPs (appearing as gaps between two words) or non-zero noun phrases (NPs, appearing as spans consisting of one word or a sequence of several words) in the novel jointly. Formally, given a piece of novel text $D = \{w_1, w_2, \dots, w_d\}$ where w_i represents the i-th word consisting of one or more tokens (a token of Chinese is always equivalent to a <u>character</u> [5]), $S = \{s_{11}, s_{12}, \dots, s_{dd}\}$ is the set of spans, where s_{ij} represents the span starting from w_i and ending at w_j, and $G = \{g_1, g_2, \dots, g_d, g_{d+1}\}$ is the set of gaps, where

[4] https://github.com/google-research/bert.

[5] Notice that underlined 'character' means a mark used in writing or printing, not a person image created by the author.

g_i represents the gap before w_i and g_{d+1} is the gap behind w_d. For simplicity, define $U = \{u_1, u_2, \ldots, u_m\}$ is the union of S and G, where u_i is a unit which can be either a span or a gap.

Our task is to find a subset $P = \{X_1, X_2, \ldots, X_z\}$ of U, where z is the number of characters that appear in D. Each set X_i refers to a character $char_i$, and any two distinct set X_i and X_j refer to different characters. Each unit in X_i is an NP or ZP referring to $char_i$. Each unit u_i will get a score for the possibility of referring to a character so that top m (a hyper-parameter) units can be selected. Then each pair of units u_i and u_j in the top-m-units will be scored for how likely they refer to a same character, so that the top m units can find their own antecedent with highest score, thereby form the character set P.

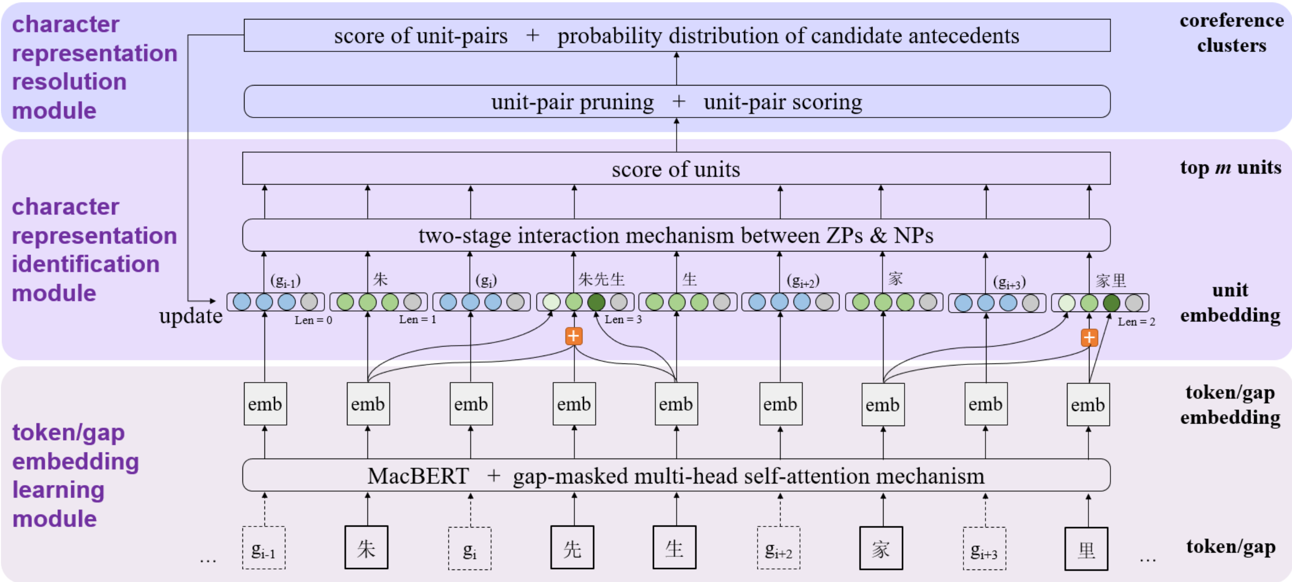


Figure 1: The architecture of ARMCC

3.2 Architecture and Training Details

ARMCC consists of three modules: token/gap embedding learning, character representation identification and character representation resolution. Its architecture is shown in Figure 1.

Token/gap embedding learning module aims to compute token embeddings and gap embeddings. It involves two steps. First, learning initial token/gap embeddings using MacBERT [6], which has been proved to achieve state-of-the-art performances on many text-level NLP tasks, such as reading comprehension and question answering (Cui et al., 2020):

$$\tilde{\boldsymbol{e}}_1, \tilde{\boldsymbol{e}}_2, \ldots, \tilde{\boldsymbol{e}}_r = MacBERT(t_1, t_2, \ldots, t_r)$$

$$\tilde{\boldsymbol{g}}_i = FFNN_g\,([\tilde{\boldsymbol{e}}_{i-1}; \tilde{\boldsymbol{e}}_i])$$

$\tilde{\boldsymbol{e}}$ and $\tilde{\boldsymbol{g}}$ are the initial token embeddings and gap embeddings respectively. t_i is the i-th token in D and tokenization is done by the WordPiece tokenizer (Wu et al., 2016). $FFNN_g$ is a feed-forward neural network. Novels are usually independent to each other and characters are usually fictional, so a same person name in two novels refers to completely unrelated characters. Learning dynamic embeddings can make sure that embeddings of person names in different novels differ from each other. It should be noted that we replace all the Chinese quotation marks in OntoNotes-*Ch* with pairwise marks in the official vocab file of BERT-Chinese model before tokenization because the most common quotation marks in Chinese are not included in the vocab file and will be counted as unknown tokens after tokenization. Dialogs make up a large part of novels and they are closely intertwined with narrative sentences. Quotation marks are vital delimiters between quotes and narrative sentences where character representations are very likely in different forms. The lack of quotation marks might cause a fatal problem for character-resolution for novels.

Second, learning final token/gap embeddings by a gap-masked multi-head self-attention mechanism:

$$\{\boldsymbol{e}, \boldsymbol{g}\} = linear_{MH}(softmax(\boldsymbol{W})\boldsymbol{V})$$

$$\boldsymbol{W}_i = \begin{cases} \dfrac{\boldsymbol{Q}\boldsymbol{K}_i^T}{\sqrt{d_k}}, & i \in D \\ -inf, & i \in G \end{cases}$$

$$\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} = linear_{q,k,v}\,(\{\tilde{\boldsymbol{e}}_1, \tilde{\boldsymbol{e}}_2, \ldots, \tilde{\boldsymbol{e}}_r, \tilde{\boldsymbol{g}}_1, \tilde{\boldsymbol{g}}_2, \ldots, \tilde{\boldsymbol{g}}_d\})$$

$\{\boldsymbol{e}, \boldsymbol{g}\}$ is the final embedding matrix of the input sequence. $linear_{MH}$ and $linear_{q,k,v}$ are four different linear functions. The multi-head self-attention mechanism is applied to capture the semantic relationship within the input sequence. The gap-masked mechanism computes embeddings for gaps from their surrounding tokens by setting the attention scores of gaps to $-inf$ when computing attention weights, thus allowing the final token embeddings and gap embeddings to be in the same space.

Character representation identification module aims to compute scores of units for their likelihood of being a ZP/NP referring to a character. Because Chinese script is not alphabetic, when splitting the text using WordPiece tokenizer, a token of Chinese is always equivalent to a character rather than a word. However, gaps between two Chinese characters inside a word cannot be a ZP or the boundary of an

[6] https://githubplus.com/ymcui/MacBERT.

NP in natural language. Therefore, we exclude units that appear inside or across a word, including gaps inside a word and spans whose boundaries are inside a word. As the example '…朱(Zhu) 先生(Mr.) 家(house) 里(in)… (… in Mr. Zhu's house …)' in Figure 1, the gap between 先 and 生 and the spans which end at 先 or start from 生 are excluded when scoring units. We count that, if such false units are not excluded, they will account for 5.79% of the predicted top-m-units.

This module involves two steps. First, learning unit embeddings following Lee et al. (2017):

$$\boldsymbol{h}_{u_i} = [\boldsymbol{start}_{u_i}; \boldsymbol{end}_{u_i}; \boldsymbol{att}_{u_i}; \phi(u_i)]$$

$$\boldsymbol{att}_{u_i} = softmax\left(FFNN_\alpha(\boldsymbol{E}_{u_i})\right)\boldsymbol{E}_{u_i}$$

For span embeddings, $\boldsymbol{start}_{u_i}$ and $\boldsymbol{end}_{u_i}$ are the final embeddings of the span's first and last token respectively, and $\boldsymbol{att}_{u_i}$ is the span's attention embedding which is computed based on all tokens within the span. For gap embeddings, $\boldsymbol{start}_{u_i}$, $\boldsymbol{end}_{u_i}$ and $\boldsymbol{att}_{u_i}$ are all the final embedding of the gap. $\phi(u_i)$ is a feature vector indicating the length of u_i. Compared with early works which encoded only the first and last tokens of a span, this method is more applicable to encode Chinese person names because names of blood brothers and sisters often differ only in the middle character, such as the four sisters (贾元春, 贾迎春, 贾探春, 贾惜春) in the Jia Mansion in *A Dream of the Red Mansions* (Cao and Gao, 1982).

Second, computing the likelihood that each unit is a ZP/NP referring to a character, using a two-stage interaction mechanism between ZPs and NPs. The positions of character representations in the form of ZPs and NPs are mutually exclusive, namely a ZP does not appear within an NP. Therefore, each unit has a relevant unit set R_{u_i} containing units that are mutually exclusive with u_i in position, and the scores of u_i and R_{u_i} should be negatively correlated. u_i is scored in two stages:

$$(1)\quad \tilde{s_u}(u_i) = \begin{cases} FFNN_m\left(\left[\boldsymbol{h}_{u_i}; \boldsymbol{h}_{R_{u_i}}\right]\right),\ u_i \in S \\ FFNN_z\left(\left[\boldsymbol{h}_{u_i}; \boldsymbol{h}_{R_{u_i}}\right]\right),\ u_i \in G \end{cases}$$

$$\boldsymbol{h}_{R_{u_i}} = softmax\left(FFNN_\beta\left(\boldsymbol{H}_{R_{u_i}}\right)\right)\boldsymbol{H}_{R_{u_i}}$$

$$(2)\quad s_u(u_i) = \tilde{s_u}(u_i) - \text{mean}\left(\cup_{u_j \in R_{u_i}} \tilde{s_u}(u_j)\right)$$

$\boldsymbol{H}_{Ru_i}$ is the matrix composed of the embeddings of all the units in R_{u_i}, and $\boldsymbol{h}_{Ru_i}$ is the attention embedding of $\boldsymbol{H}_{Ru_i}$. Stage 1 scores each unit based on the embeddings of its own and relevent units. Stage 2 further modifies each unit-score by scores of its relevent units. The top m units can be selected according to $s_u(u_i)$. CHEN-Model ignores the last tokens of all sentences when scoring units, which hardly affects the model's performance because the end token of a sentence is usually a punctuation. However, novel writers are not always used to adding end-of-sentence punctuations, especially in manuscripts, largely due to high degrees of freedom in writing. We therefore fix this problem in ARMCC.

Character representation resolution module aims to compute scores of unit-pairs for their likelihood of referring to a same character, and is applied only to the top m units. In order to reduce the computational complexity, the $m \times m$ unit-pairs are pruned following Lee et al. (2018) before scoring, and k candidate antecedents are reserved for each unit. Then the likelihood of referring to a same character for the remaining $m \times k$ unit-pairs are computed:

$$s(u_i, u_j) = s_u(u_i) + s_u(u_i) + \tilde{s_a}(u_i, u_j) + s_a(u_i, u_j)$$

$$\tilde{s_a}(u_i, u_j) = linear_{coarse}(\boldsymbol{h}_{u_i})\boldsymbol{h}_{u_j}$$

$$s_a(u_i, u_j) = FFNN_a\left(\left[\boldsymbol{h}_{u_i}; \boldsymbol{h}_{u_j}; \boldsymbol{h}_{u_i} \cdot \boldsymbol{h}_{u_j}; \delta(u_i, u_j)\right]\right)$$

$\tilde{s_a}(u_i, u_j)$ is the initial score of the unit-pair calculated when pruning unit-pairs. The unit-pair embeddings inputted to $s_a(u_i, u_j)$ consist of four parts: the unit embeddings of u_i and u_j respectively, the cosine similarity between them, and a feature vector $\delta(u_i, u_j)$ that indicates the distance between them and whether they are spoken by a same person. We encode speaker information as a feature because quotes take a large proportion of novels and our experiments show that speaker information is significantly helpful for resolving character representations in conversational texts (see details in Section 4.1). The final unit-pair-score $s(u_i, u_j)$ indicates the likelihood that both u_i and u_j refer to a character and the characters they refer to are the same. The possibility distribution of the candidate antecedents can be calculated based on the scores of unit-pairs and be used to update unit embeddings following Lee et al. (2018):

$$\boldsymbol{h}_{u_i}^{n+1} = \boldsymbol{f}_{u_i}^n \circ \boldsymbol{h}_{u_i}^n + (\boldsymbol{1} - \boldsymbol{f}_{u_i}^n) \circ \boldsymbol{a}_{u_i}^n$$

$$\boldsymbol{f}_{u_i}^n = \sigma\left(FFNN_f([\boldsymbol{h}_{u_i}^n; \boldsymbol{a}_{u_i}^n])\right)$$

$$\boldsymbol{a}_{u_i} = \sum_{u_j \in U_i} softmax\left(s(u_i, u_j)\right)\boldsymbol{h}_{u_i}$$

$\boldsymbol{h}_{u_i}^n$ is the n-th update result of $\boldsymbol{h}_{u_i}$, $\boldsymbol{a}_{u_i}$ is the attention embedding of $\boldsymbol{h}_{u_i}$, and $\boldsymbol{f}_{u_i}^n$ is a gate vector used to control whether the updated unit embedding retains more information in $\boldsymbol{h}_{u_i}$ or $\boldsymbol{a}_{u_i}$ during the n-th update. The update to unit embeddings enables the possibility distribution of the candidate antecedents to be adjusted based on the predicted coreference clusters during model training.

The **cost function** of ARMCC is combined with identification loss and resolution loss following Zhang et al. (2018):

$$L_{CCARM} = \lambda \sum_{u_i \in U} L_{identify}(u_i) + \sum_{u_i \in U'} L_{resolve}(u_i)$$

$$L_{identify}(u_i) = -y_i \log(\hat{y}_i) - (1 - y_i)\log(1 - \hat{y}_i)$$

$$L_{resolve}(u_i) = -\log \sum_{u \in U_i \cap GOLD(u_i)} P(u)$$

The identification loss is applied to all the units which do not appear inside or across a word, and y_i and $\hat{y}_i$ are the predicted possibility and gold label (0/1) of u_i referring to a character respectively. The resolution loss is applied to the top m units and $P(u)$ is the predicted possibility of u being the antecedent for u_i. U_i and $GOLD(u_i)$ are the set of candidate and gold antecedents for u_i respectively.

Because ARMCC focuses on character representations, we build a new dataset from OntoNotes-*Ch* for training. We delete sections and coreference clusters which do not involve person names, and then randomly split it into training/ development/test datasets by 8:1:1. Accordingly, we decrease the hyper-parameter m from half of the number of tokens to 20%.

Overall, the modifications by ARMCC on the basis of CHEN-Model are as follows: we filter units according to word boundaries, restore the delimitation function of Chinese quotation marks, use MacBERT which is better at text-level tasks to learn initial embeddings of input text, fix the problem of ignoring the last tokens of all sentences and build a new dataset focusing on characters from OntoNotes-*Ch* for training.

3.3 Experiments and Results

Table 1 shows the experimental results of ARMCC. It achieves 70.62% F_1-score for non-zero coreference resolution and 32.25% for zero pronoun resolution for character representations on the development set, suggesting that it can effectively resolve most non-zero NPs and part of ZPs referring to characters.

To validate the efficiency of the approaches including encoding speaker information, filtering units according to word boundaries, replacing quotation marks and the pre-trained model, we also perform ablation studies (see results in Table 2). It shows that these approaches are all beneficial for improving ARMCC's performance. Word boundaries and MacBERT make greater contributions, especially in zero pronoun resolution. Since the measures of encoding speaker information and replacing quotation marks are specially implemented for Chinese novels, we believe that the superiority will be greater when using ARMCC on Chinese novels.

Non-zero coreference resolution										Zero pronoun resolution		
MUC			B^3			$CEAF_{\phi4}$			Avg.F_1	P	R	F_1
P	R	F_1	P	R	F_1	P	R	F_1				
77.75 (69.21)	64.76 (67.51)	70.67 (68.35)	70.24 (67.08)	59.20 (64.09)	64.25 (65.55)	77.29 (77.39)	76.68 (75.38)	76.98 (76.37)	**70.62** (70.09)	37.55 (26.76)	28.26 (28.50)	**32.25** (27.60)

※ The three metrics (%) of non-zero coreference resolution are MUC (Vilain et al., 1995), B^3 (Bagga and Baldwin, 1998) and $CEAF_{\phi4}$ (Luo, 2005). The metric (%) of zero pronoun resolution follows Zhao and Ng (2007). Figures outside brackets are values on the development set and those in brackets are values on the test set.

Table 1: Results of ARMCC on the new dataset

Model	Non-zero coreference resolution				Zero pronoun resolution
	MUC.F_1	B^3.F_1	$CEAF_{\phi4}$.F_1	Avg.F_1	F_1
ARMCC	70.67	**64.25**	**76.98**	**70.62**	**32.25**
(-) speaker information	**70.79**	62.89	75.82	69.83	30.38
(-) word boundaries	69.59	60.62	74.39	68.20	27.35
(-) quotation marks	70.42	63.10	75.94	69.82	30.57
(-) MacBERT	69.91	60.53	75.63	68.69	27.37

※ '(-)' represents that the information is not used or considered. '(-) MacBERT' represents changing the pre-trained model back to Chinese-BERT-base.

Table 2: Results of ablation studies

4. Speaker Extraction Model SECN

4.1 Reason for Training SECN

Dialogs are a significant component of novels, both in terms of the proportion of text and the role in storytelling. There are often a variety of character representations in quotes, as well as frequent phenomena such as turn-taking and the change of personal pronouns, which is a big challenge to anaphora resolution models. To investigate whether speaker information is helpful for Chinese anaphora resolution, we conduct several groups of comparative experiments using CHEN-Model on OntoNotes-*Ch*. There are six genres in OntoNotes-*Ch*: broadcast conversation, broadcast news, magazine, newswire, telephone conversation and web data, where the proportions of dialogs are considerably different. We train CHEN-Model with and without encoding the information of genres and speakers when computing scores of unit-pairs for how likely they refer to a same entity, and test it on the whole dataset and datasets divided into genres. The results are shown in Table 3. When using genre and speaker information, CHEN-Model performs better on genres where dialogs account for a large proportion of the contents. This is particularly noticeable on broadcast conversation and telephone conversation (with grey shading), whose F_1-score for mixed anaphora resolution increase by about 3%. Therefore, we conclude that speaker information is helpful for resolving character representations in Chinese novels and use speaker information in ARMCC. Speakers of each sentence have been annotated in OntoNotes-*Ch*, but novels lack annotated speaker information that can be provided to ARMCC, which motivate us to train a speaker extraction model for automatically extracting speakers of quotes for Chinese novels. Moreover, the speaker extraction model can also be solely used for producing audiobooks and radio drama scripts and constructing dialog networks for novels.

4.2 Training and Dataset Reconstruction

The common NLP task associated with speakers is speaker identification (or quote attribution), whose routine scheme is a two-step procedure: extracting the name of the speaker from the context of the quote and then link the name to a character entity. The latter step can be seen a subtask of anaphora resolution, which can be conducted by ARMCC. Therefore, SECN focus on the former step, which we called speaker extraction, aiming to extract a single span from a piece of novel text containing a quote to be the speaker of the quote.

Span prediction (or extractive question answering) is a kind of reading comprehension tasks: given a passage and a question about the passage, the goal is to extract a single span in the passage to be the answer to the question. Fine-tuning pre-trained language models such as SQuAD BERT model (Devlin et al., 2019) has achieved enormous successes in reading comprehension tasks. Inspired by these successes, we treat speaker extraction as a span prediction task and train SECN based on SQuAD v1.1 BERT model. Specifically, given a piece of novel text containing a quote and a question 'who is the speaker of the quote?', SECN needs to extract a span from the text to be the speaker of the quote, without being given any extra information. The inputs are in the form of a concatenated sequence '[CLS] + question + [SEP] + context + [SEP]', which will be truncated or padded with padding tokens to a fixed length. '[CLS]' is a classifier token and '[SEP]' is a sentence separator token. SECN predict the start and end positions of the answer spans, and thus output the answer spans. Keeping in line with ARMCC, we use MacBERT to learn token embeddings. For lack of space, we present the architecture of SECN in Appendix A.

Dataset	$F1_{mixed}$		$F1_{non\text{-}zero}$		$F1_{zero}$	
	w/o features	w/ features	w/o features	w/ features	w/o features	w/ features
all	61.94	62.61	66.58	67.30	33.04	33.38
broadcast conversation	52.42	55.07	58.03	61.09	23.46	34.47
broadcast news	71.59	72.58	75.56	76.73	37.28	41.75
magazine	49.81	47.85	53.42	51.77	12.44	16.88
newswire	72.27	71.92	74.00	74.02	25.84	24.87
telephone conversation	57.40	60.44	65.42	70.10	35.36	37.53
web data	50.16	50.80	55.19	56.61	25.20	27.40

※ The features include genre number and speaker information (whether the two units are from the same speaker). The F_1-score (%) for non-zero coreference resolution and mixed anaphora resolution is an average of MUC, B^3 and $CEAF_{\phi 4}$.

Table 3: Results of CHEN-Model with and without encoding features for different genres in OntoNotes-*Ch*

```
{"data":
    [{"title": "novel title",
      "paragraphs":
          [{"context": preceding five sentences + quote + next five sentences,
            "qas": [{"question": who is the speaker of the quote?,
                     "answers": [{"answer_start": position of the first token of the speaker in the text,
                                  "text": the speaker (a span in the context)}],
                     "id": a serial number between 0 and 30,335 }]}]}]}
```

Figure 2: Instance template of $\mathbb{D}_{SECN}$

As a supervised deep-learning method, fine-tuning pre-trained models also requires training datasets, so we integrate the corpora built in previous studies and convert them into SQuAD-format. Jia et al. (2021) manually annotated a corpus ($\mathbb{D}_{JY}$ for short) [7] containing 28,597 quotes from Jin Yong's three novels: *Legend of the Condor Heroes* (Jin, 1994), *Return of the Condor Heroes* (Jin, 1994), and *Heaven Sword and Dragon Sabre* (Jin, 1994). The instance format is '{preceding five sentences, quote, next five sentences, list of candidate characters, gold speaker, start position of gold speaker, end position of gold speaker}'. The list of candidate characters denotes the characters appearing in the context of altogether eleven sentences. Chen et al. (2019) manually constructed a global character list containing all characters and their aliases for a famous Chinese novel *World of Plainness* (Lu, 1986-1988), and built a corpus ($\mathbb{D}_{WoP}$ for short) [8] containing 1,835 quotes from the novel. The instance format is '{preceding ten sentences, quote, next ten sentences, index list of candidate characters, gold speaker index}'. It is notable that Jin Yong's novels and *World of Plainness* vary in the forms of person names in narrative sentences. In Jin Yong's novels, they are always full names, such as 穆念慈 (Mu Nian-ci) and 欧阳锋 (Ou-Yang Feng). But in *World of Plainness*, they are often abbreviated to be aliases, for example, the family name 孙 (Sun) of 孙少平 (Sun Shao-ping) is often omitted. Considering that name abbreviations are common in novels, we incorporate instances in $\mathbb{D}_{WoP}$ as well as $\mathbb{D}_{JY}$ into our dataset.

We integrate $\mathbb{D}_{JY}$ and $\mathbb{D}_{WoP}$ and convert them into the JSON format of SQuAD (Rajpurkar et al., 2016), and call it $\mathbb{D}_{SECN}$ for short. The instance format of SQuAD is '{context, {question, {answer, start position of answer}}, instance index}'. The answer is a span extracted from the context. To maintain consistency throughout the whole dataset, we concatenate the preceding five sentences, quote and the next five sentences into a context. The format of questions is '……的说话人是谁？(who is the speaker of … ?)', and the answer is the gold speaker of the quote. $\mathbb{D}_{JY}$ has provided the start positions of gold speakers, while $\mathbb{D}_{WoP}$ labels gold speakers with the character indices, each of which corresponds to a character's full name and aliases, such as 孙玉厚 (Sun Yu-hou), 玉厚 (Yu-hou), 厚 (Hou). In order to convert the gold character-indices to gold speakers, we simply select the nearest full name of each quote to be its gold speaker. If the

[7] https://github.com/huayi-dou/The-speaker-identification-corpus-of-Jin-Yong-novels. $\mathbb{D}_{JY}$ contains quotes from Jin Yong's three novels, but only one is reported in their paper.

[8] https://github.com/chenjiaxiang/Chinese-dataset-for-speaker-identification. We count that $\mathbb{D}_{WoP}$ contains 1,835 quotes, rather than 2,548 reported in their paper.

corresponding full name cannot be found, we then select its first alias, then the second alias, and so on. Removing quotes whose speakers are out of the preceding and next five sentences, we reserve 1,739 instances in $\mathbb{D}_{WoP}$. The ultimately integrated dataset $\mathbb{D}_{SECN}$ of SQuAD-format contains 30,336 quotes. Figure 2 shows its instance template. Additionally, in order to give full play to the positive role of quotation marks, we replace Chinese quotation marks in $\mathbb{D}_{SECN}$ with the same pairwise marks mentioned in ARMCC.

4.3 Experiments and Error Analysis

Experimental results are shown in Table 4. We cannot compare the performances of SECN with models proposed by Chen et al. (2019) and Jia et al. (2020) because their experimental results are difficult to integrate, and experiments in Jia et al. (2021) are conducted on only two simple categories of speakers (person names and pronouns). Therefore, we propose two baselines. **Baseline1**: we extract the person name nearest to the quote in the whole context as the predicted speaker, because the speaker is usually mentioned near the quote to avoid confusion in dialogs. **Baseline2**: we extract the person name nearest to the quote in the preceding five sentences as the predicted speaker, because the speaker usually occurs before the quote.

	$EM_{random\text{-}10\%}$	$EM_{new\text{-}novel}$
Baseline1	29.04 (30.98)	30.15 (29.68)
Baseline2	81.64 (83.49)	86.55 (85.61)
SECN	**96.80** (96.80)	**94.92** (94.62)

※ Random-10% denotes the development/test sets consisting of 10% instances randomly selected. New-novel denotes the development/test sets consisting of half of the instances from *Heaven Sword and Dragon Sabre*.

Table 4: Results of SECN for Chinese novels

We set up training, development and test datasets in two different methods to evaluate SECN. Firstly, we randomly split $\mathbb{D}_{SECN}$ into training/development/test sets by 8:1:1. The exact match (EM) accuracy of SECN is up to 96.80%, not only higher than the two baselines (29.04% and 81.64% on the development set), but even higher than the consistency rates of annotators (93.95% for $\mathbb{D}_{JY}$ and 94.44% for $\mathbb{D}_{WoP}$). Considering that fictional characters in different novels are independent of each other, we also randomly and averagely divide instances from *Heaven Sword and Dragon Sabre* into the development and test sets, while regard instances from the other three novels as the training set. Experimental results show that the EM accuracy of SECN is 94.92% on the development set, just a little lower than that in the random-10% set. We believe that SECN's performance on a new novel can be improved by annotating a small amount of data and using it for fine-tuning. We also perform a quantitative error analysis and summarize four main error types as follows, which can serve as a reference for manually correcting the results extracted automatically by SECN. For lack of space, we only present the data and analysis in the main body. The complete examples and their translations can be found in Appendix B.

(1) The context is too complicated for models to predict speakers. This kind of errors are the most common errors, accounting for 44.33% of errors. On the one hand, the storyline is sometimes very complex, such as the context in Example 3-1. The two characters, 黄药师 (Huang Yao-shi) and 杨过 (Yang Guo), haven't shown up when the quote 'Yes!' is spoken out. Correct answer can be predicted only if SECN correctly encodes the semantics that the 'old voice' summons Yang Guo and Yang Guo replies 'Yes!' to him. On the other hand, writers sometimes omit the prompts of speakers in the scene where few people are involved in the dialog. It is a common omission type in *World of Plainness*, such as Example 3-2. There are only two characters, 少平 (Shao-ping) and 金波 (Jin Bo), speaking to each other, and the author directly narrated their quotes and movements, without any speech act verb. In this case, readers might easily distinguish the speakers of quotes by comprehensively analyzing what they say, but it is very difficult for the model.

(2) Gold speaker is far from the quote, while the predicted speaker is closed to the quote. This is also a common error type, accounting for 17.53% of errors. As shown in Example 3-3, the predicted speaker 郭靖 (Guo Jing) is in the pre-sentence next to the quote, while the gold speaker 郝大通 (Hao Da-tong) is the first word of the whole context, much further than Guo Jing. Actually, in cases like Example 3-3, there are usually pronouns or zero pronouns closer to the quote referring to the gold speaker, such as the pronoun 他 (he) (in bold) near the quote which refers to Hao Da-tong. They might be resolved by the character resolution model.

(3) The answer is almost correctly located, but there is a small deviation between the start or end position of the predicted and gold speakers. The percentage of this type of errors is 13.40%. As shown in Example 3-4, the gold two-tokens speaker 天鸣 (TianMing) is the front part of the predicted four-tokens speaker 天鸣禅师 (abbot TianMing), which is a NP in the form of 'name + capacity'. That is to say, the answer's start position is correctly predicted, but the predicted end position is two tokens after the gold end position. Although the extracted speaker is not completely correct, SECN extracts a part of the correct speaker. This type of errors often occurs in cases of 'name + capacity' NPs, compound surnames which are special in Chinese, e.g. 公孙 (Gong-Sun), and atypical names, e.g. 赤老温 (Tchila'un). These errors may be avoided using predefined special names and rule-based methods.

(4) The information provided by the context is inadequate to extract a correct speaker, even if annotated manually. The percentage of this type of errors is 7.22%. Each context in $\mathbb{D}_{SECN}$ contains only eleven sentences, which is a very small part of the whole novel, so an annotator can't annotate the correct speaker if the information about the speaker is provided outside the context. For instance, the context in Example 3-5 provides no information about who is 大哥 (Big Brother) (in bold) of the 'Six Freaks', so SECN is unable to determine whether Big Brother corresponds to 全金发 (Quan Jin-fa) or 柯镇恶 (Ke

Zhen'e). This error type is inevitable for speaker extraction models and can only be corrected manually.

5. Experiments and Result Analysis on a Specific Chinese Novel

Because there are no novels in Ontonotes-*Ch*, the experimental data is insufficient to demonstrate the effectiveness of our approaches on Chinese novels. We test our approaches on *White Deer Plain*, a representative local novel in Chinese contemporary literature written by Chen Zhong-shi, and make a detailed analysis on resolution results. We first show the preliminaries and experiments, then analyze the results of character resolution, including both detailed analysis of several passages and comprehensive analysis about difficulties and challenges, and propose potential solutions to some problems.

5.1 Preliminaries and Experiments

Firstly, following Chen et al. (2019), we divide the raw novel texts into quotes and narrative sentences according to colons and quotation marks, and then divide narrative sentences according to end-of-sentence punctuations. We also replace Chinese quotation marks with the same pairwise marks mentioned above. Secondly, we prepare the SQuAD-format test file for SECN and predict speakers for each quote. Thirdly, we segment the words using Language Technology Platform (LTP) (Che et al., 2010), which are used in MacBERT to identify the boundaries of Chinese words, and prepare the test file containing speaker information for ARMCC. Different from genres in Ontonotes-*Ch*, there are always different characters appearing intensively in novels. Inputting long passages from novels would be confusing for the model, thus we limit the length of each input passage to three pieces, with no more than 128 tokens per piece. Finally, we resolve character representations in *White Deer Plain* using ARMCC that performs well on both non-zero coreference resolution (F_1-score = 68.58%) and zero pronoun resolution (F_1-score = 32.03%).

5.2 Test Result Analysis and Challenge Summary

We recognize 7,953 occurrences of person names in *White Deer Plain* using the named entity recognition module of LTP, while ARMCC detects 19,794 occurrences of character representations (including 4,647 ZPs), which is one and a half times more than that of person names. That is to say, by resolving different representations of characters, we can preserve much more information about characters than focusing only on person names. This is very beneficial for many higher-level automatic tasks on characters such as interaction network construction, plot summarization and similar character recommendation.

Following is analysis of the resolution results. The complete examples in case analysis and their translations can be found in Appendix C.

Case Analysis. We analyze several character-resolution results in detail to investigate the abilities and problems of our approaches. In Example 4-1, most zero and non-zero pronouns referring to the protagonist 嘉轩 (Jia-xuan) both within a sentence and across sentences are successfully resolved, including a third-person pronoun 他 (he) in a narrative sentence, a first-person pronoun 我 (I) spoken by himself and several second-person pronouns 你 (you) spoken by others in the quotes, a reflexive pronoun 自己 (himself), and a dropped subject in a verbal phrase. Jia-xuan can be further resolved by his full name 白嘉轩 (Bai Jia-xuan) in another passage that contains both Jia-xuan and Bai Jia-xuan. ARMCC is also able to resolve NPs referring to a characters that are even not clearly named. For example, 三官庙老和尚 (the old monk in the Sanguan Temple), 年逾六旬的老和尚 (the sexagenarian old monk) and 老和尚 (the old monk) in Example 4-2 refer to a same monk. Moreover, ARMCC correctly resolve several appositives referring to different characters without any confusion in Example 4-3, such as 一营长 (the commander of the first battalion) referring to 白孝文 (Bai Xiao-wen), 二营长 (the commander of the second battalion) referring to 焦振国 (Jiao Zhen-guo) and 县党部书记 (the Party branch secretary of the county) referring to 岳维山 (Yue Wei-shan). This means it can also be used to extract characters' attributes from text. However, as we can see in the above examples, there are still many ZPs that have not been correctly identified and resolved, such as the dropped subjects of 在雪地里撒尿 (pee in the snow), 发现那一坨无雪的慢坡地 (find the irrigated farmlands without snow), 挖出怪物 (dig out the monster) and 拉屎伪造现场 (shit and forge the scene) in Example 4-1. Admittedly, there is still much room for improvement in ARMCC's performance on zero pronoun resolution. The lower F_1-score for zero pronoun resolution also supports this problem. This problem can be temporarily alleviated by using dependency parsers. In a dependency tree, if a word is the parent node of a character representation outside its clause, there is likely to be a ZP referring to the character inside its clause. Additionally, ARMCC might break down when multiple character representations referring to different characters occur together, such as Example 4-4. It fails to distinguish 鹿三 (Lu San), 嘉轩(Jia-xuan), 白赵氏 (Mrs. Bai) and 秉德老汉 (Old Man Bing-de), but gathers them into a same cluster. Example 4-5 shows another case where personal pronouns, nicknames, and other representations occur intensively and turn-taking is frequent, although they refer to only two different characters, 黑娃 (Little Black) and 兆鹏(Zhao-peng). It fails to distinguish these various representations and gathers them into a same cluster. These errors may be caused by the inappropriate m and k, namely the numbers of units and unit-pairs selected by the model. When we reduce m and k for such passages, the problem can be alleviated, but some correct units will be missed at the same time. We suggest that this kind of errors may be relieved by using the list of character names and simple rules of splitting clusters. It may also be useful to set m and k dynamically according to the number of different characters.

In summary, with assistance of SECN, ARMCC is able to resolve most personal pronouns, aliases, NPs and ZPs referring to characters in Chinese novels, so it can adequately satisfy the needs of coarse-grained researches on characters in Chinese novels. But these approaches are not yet sufficient for fine-grained researches without manual assistance. The character resolution results can be further optimized by combining with syntactic parsers and simple heuristic rules.

Overall Analysis. In addition to case analysis as above, we observe almost all the resolution results and summarize other major hurdles and challenges to be overcome for the character-resolution for Chinese novels. Firstly, there are always large number of plots describing scenes of chat and talk which involve or focus on only two characters, and therefore first-person and second-person pronouns change frequently with the shift of perspectives. Some writers' preferences of omitting prompts of speakers further bring more difficulties. We suggest that it may be useful to perform a pre-resolution of first-person and second-person pronouns in the singular form using speakers extracted by SECN, namely resolving first-person singular pronouns in quotes to their speakers and resolving second-person singular pronouns to a different speaker of the nearest quotes in advance.

Secondly, there are a variety of appellations in Chinese, which are respectively suitable for different status, occasions, relationships, occupation, and so on, and novels contain much more appellations than other genres. We believe that a complete collection of appellations and their rules is necessary and must be useful for character resolution for Chinese novels.

Thirdly, it is a common phenomenon in novels that character representations are not one-to-one referred. However, mainstream methodologies of anaphora resolution aggregate mentions and ZPs into non-intersecting clusters, so they are unable to deal with complex coreference relationships. For example, 鹿家父子 (the father and son of Lu family) should be resolved to 鹿泰恒 (Lu Tai-heng) and 鹿子霖 (Lu Zi-lin), while 白鹿两家的主事者 (heads of Bai family and Lu family) should be resolved to Bai Jia-xuan and Lu Zi-lin. This problem needs to be taken seriously when designing character resolution models in future.

Fourthly, too few novels are included in the annotated datasets commonly used to train anaphora resolution models, and even fewer Chinese novels. The characteristics of novels limit the performance of domain-general models on novels because they cannot learn these characteristics from other genres. Novel writers have much freedom in writing and the contents are not limited by time, setting, language and even reality. For example, as a local novel, although *White Deer Plain* was written in modern Chinese, there are still many dialect words in the quotes, such as 大 (dad) which means big in Mandarin. ARMCC almost completely failed to resolve these words. We believe it is necessary to make great efforts to build novel-specific annotated resources.

Fifthly, character-resolution for plots about hidden identity, including recognizing the pretenders, tracing the criminals, distinguishing the substitute and the real one and those even against the reality, can hardly demonstrate good performance without manual assistance. For example, a plot in *White Deer Plain* describes that Lu San's body were possessed with the ghost of the deceased 田小娥 (Tian Xiao'e), so they are hard to distinguish. Effective methods to overcome this hurdle need to be proposed.

6. Conclusions

This paper presents our approaches to resolving different representations of characters for Chinese novels, which is an exploration of adapting NLP technologies to literary-specific tasks. We make some modifications on a state-of-the-art anaphora resolution model to design a character resolution model for Chinese novels ARMCC, according to the text characteristics and application requirements. It achieves 70.62% F_1-score for non-zero coreference resolution and 32.25% F_1-score for zero pronoun resolution for character representations. Because speaker information of quotes can effectively optimize the character resolution model, we also train a widely used BERT fine-tuned model for speaker extraction SECN on a reconstructed dataset of Chinese novels. Its EM accuracy is up to 96.80% and it does not require any extra information apart from the raw novel texts. We test our approaches on a Chinese novel *White Deer Plain*. Case analysis of the resolution results shows that, with assistance of SECN, ARMCC can satisfy the needs of coarse-grained researches and greatly reduce human effort in fine-grained researches on characters in Chinese novels. But overall analysis of the resolution results shows that there are also many problems to be overcome. We summarize main hurdles and challenges for character-resolution in Chinese novels and propose some potential solutions to some problems.

7. Acknowledgments

We thank the reviewers for their helpful comments. This work is supported by Major Program of National Philosophy and Social Science Foundation (No. 18ZDA238) and Major Commissioned Project of National Language Commission (No. ZDA145-6).

8. Bibliographical References

Aloraini, A. and Poesio, M. (2020). Anaphoric zero pronoun identification: a multilingual approach. In *Proceedings of the 3rd Workshop on Computational Models of Reference, Anaphora and Coreference*, 22-32.

Bagga, A. and Baldwin, B. (1998). Algorithms for scoring coreference chains. In *Proceedings of the 1st International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, 1: 563-566.

Cao, X. and Gao, E. (1982). *A Dream of the Red Mansions*. Beijing: People's Literature Publishing House.

Chang, T., Lv, S., Wang, X., and Wang, D. (2017). Zero pronoun identification in Chinese language with deep neural networks. In *Proceedings of the 2nd International Conference on Control, Automation and Artificial Intelligence*, 518-522.

Che, W., Li, Z., and Liu, T. (2010). LTP: a Chinese language technology platform. In *Proceedings of t he 23rd International Conference on Computational Linguistics: Demonstrations*, 13-16.

Chen, C. and Ng, V. (2016). Chinese zero pronoun resolution with deep neural networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1: 778-788.

Chen, J., Ling, Z., and Dai, L. (2019). A Chinese dataset for identifying speakers in novels. *INTERSPEECH*, 1561-1565.

Chen, S., Gu, B., Qu, J., Li, Z., Liu, A., Zhao, L., and Chen, Z. (2021). Tackling zero pronoun resolution and non-zero coreference resolution jointly. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, 518-527.

Chen, Z. (1993). *White Deer Plain*. Beijing: People's Literature Publishing House.

Cuesta-Lazaro, C., Prasad, A., and Wood, T. (2022). What does the sea say to the shore? A BERT based DST style approach for speaker to dialogue attribution in novels. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 1: 5820-5829.

Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., and Hu, G. (2020). Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, 657-668.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 4171-4186.

Iwata, S., Watanabe, T., and Nagata, M. (2021). Zero pronouns identification based on span prediction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 331-336.

Jia, Y., Dou, H., Cao, S., and Zan, H. (2020). Speaker identification and its application to social network construction for Chinese novels. *International Journal of Asian Language Processing*, 30(04): 2050018.

Jin, Y. (1994). *Heaven Sword and Dragon Sabre*. Beijing: SDX Joint Publishing Company.

Jin, Y. (1994). *Legend of the Condor Heroes*. Beijing: SDX Joint Publishing Company.

Jin, Y. (1994). *Return of the Condor Heroes*. Beijing: SDX Joint Publishing Company.

Kantor, B. and Globerson, A. (2019). Coreference resolution with entity equalization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 673-677.

Lee, K., He, L., Lewis, M., and Zettlemoyer, L. (2017). End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*: 188-197.

Lee, K., He, L., and Zettlemoyer, L. (2018). Higher-order coreference resolution with coarse-to-fine inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, 2: 687-692.

Lin, P. and Yang, M. (2020). Hierarchical attention network with pairwise loss for Chinese zero pronoun resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05): 8352-8359.

Lu, Y. (1986). *World of Plainness (Volume 1)*. Beijing: China Federation of Literary and Art Circles Press.

Lu, Y. (1987). *World of Plainness (Volume 2)*. Beijing: China Federation of Literary and Art Circles Press.

Lu, Y. (1988). *World of Plainness (Volume 3)*. Beijing: China Federation of Literary and Art Circles Press.

Luo, X. (2005). On coreference resolution performance metric. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 25-32.

Muzny, G., Fang, M., Chang, A., and Jurafsky, D. (2017). A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 1: 460-470.

Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2383-2392.

Song, L., Xu, K., Zhang, Y., Chen, J., and Yu, D. (2020). ZPR2: joint zero pronoun recovery and resolution using multi-task learning and BERT. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5429-5434.

Vilain, M., Burger, J. D., Aberdeen, J., Connolly, D., and Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Conference on Message Understanding*, 45-52.

Wang, Z., Wang, Y., Liu, T., Li, W., and Liao, J. Rule-BertAtten based quote attribution for Chinese novels. *Journal of Chinese Information Processing*, 36(3): 82-90.

Wu, W., Wang, F., Yuan, A., Wu, F., and Li, J. (2020). CorefQA: coreference resolution as query-based span prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6953-6963.

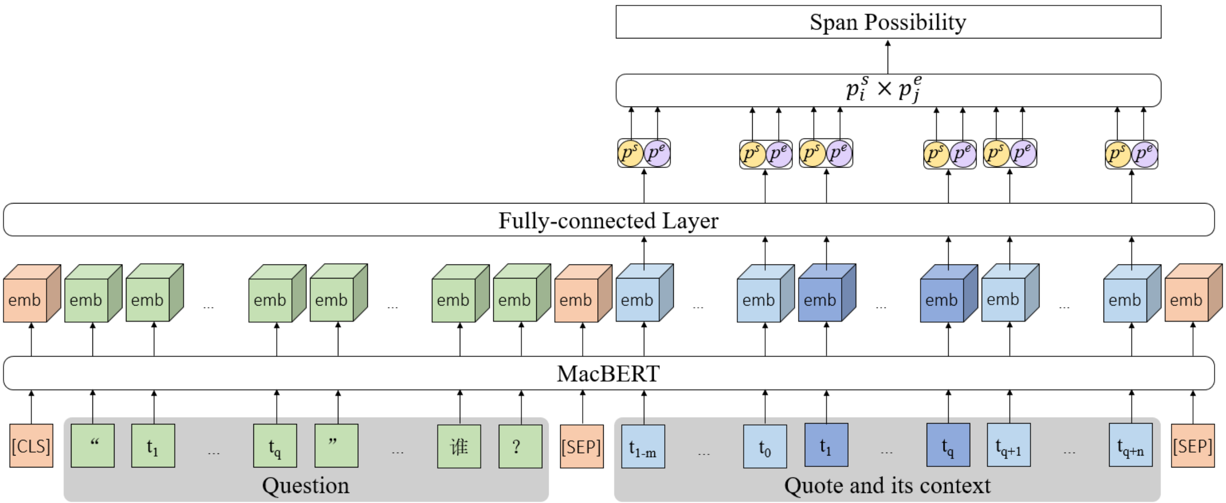
Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. (2016). Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv: 1609.08144*.

Yin, Q., Zhang, W., Zhang, Y., and Liu, T. (2017). A deep neural network for Chinese zero pronoun resolution. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 3322-3328.

Zhang, R., dos Santos, C., Yasunaga, M., Xiang, B., and Radev, D. (2018). Neural coreference resolution with deep biaffine attention by joint mention detection and mention clustering. In *Proceedings of the 56th Annual Meeting of the*

Association for Computational Linguistics, 2: 102-107.
Zhao, S. and Ng, H. T. (2007). Identification and resolution of Chinese zero pronouns: A Machine Learning Approach. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 541-550.

Appendix A: The architecture of SECN



Appendix B: Examples for Error Analysis of SECN in Section 4

In each example shown below, the quote is marked by a wavy underline, and the gold and predicted speaker are marked by a box and grey shading respectively. For brevity, we omit some extraneous spans by ellipses.

[Example 3-1] from *Return of the Condor Heroes*

黄蓉听了这暗器的破空之声，…… 惊喜之下不暇细想，纵声叫道，“是爹爹驾临么？”只听得左边旗斗中一个苍老的声音哈哈大笑，说道：“杨过小友，咱们一起下去罢！”右边旗斗中一人应声：“是！”两边旗斗之中各自跃下一人。…… 正是黄药师和杨过。……

When Huang Rong heard the projectiles hurtling through the air, …… Surprised, she did not think clearly, so she said, “Father, is that you?” An old voice from atop one of the flag poles laughed, “My friend Yang Guo, let’s go down!” The voice atop the other flag pole replied, “Yes!” and the two people jumped down together. …… It was indeed Huang Yao-shi and Yang Guo. ……

[Example 3-2] from *World of Plainness*

“你来我太高兴了！我早听说你已经不教书……我也想过，你不会死守在双水村！”“你也吃！”少平端起一大碗面片，先把一颗鸡蛋扒拉在嘴边。“我吃过了。”金波坐在一边开始抽烟，满意地看着少平吃得狼吞虎咽。“我大概吃不了这么多……”“我知道你的饭量哩！”少平噙一嘴饭，笑了。……

“I’m really glad that you are here! I have heard that you are not a teacher any more... I know that you won’t be stuck in Duaqua village!” “Let’s eat together!” Shao-ping picked up the big bowl of wheaten flakes, and pulled one boiled egg to his lips at first. “I have dined already,” Jin Bo started to smoke, satisfied with watching Shao-ping’s devouring. “Maybe that is too much for me...” “I know your appetite!” Shao-ping smiled with a mouthful of food. ……

[Example 3-3] from *Return of the Condor Heroes*

郝大通是全真教高士，道学武功，俱已修到甚高境界，…… 众人只道他要剑刺杨过，郭靖踏上一步，欲待相护，岂知**他**倒转长剑，将剑柄向杨过递去，说道：“不错，我是杀错了人。你跟孙婆婆报仇罢，我决不还手就是。”众人见他如此，无不大为惊讶。……

Hao Da-tong was an eminent Taoist of Quanzhen, he had learned martial arts, and he had reached a deep level in both areas. …… Everyone thought that he wanted to stab Yang Guo with the sword. Guo Jing took a step forward to protect Yang Guo but no one knew that **he** would turn the long sword around with the handle facing Yang Guo and say, “Correct. I killed the wrong person. You take revenge for Grandma Sun. I won’t retaliate.” When everyone saw him do this, they were surprised. ……

[Example 3-4] from *Heaven Sword and Dragon Sabre*

…… 老方丈天鸣禅师见到何足道和郭襄，合十说道：“这一位想是号称琴剑棋三圣的何居士了。老僧未能远迎，还乞恕罪。”何足道躬身行礼，说道：“晚生何足道，‘三圣’狂名，何足道哉！……”……

…… When abbot TianMing saw He Zu-dao and Guo Xiang, he put his palms together and said: “This must be the saints of chess, swordsmanship and zither, Benefactor He. Forgive us for a late welcome.” He Zu-dao returned respects and said: “My name is indeed He Zu-dao, the nickname of 3 saints is not worth mentioning. ……” ……

[Example 3-5] from *Legend of the Condor Heroes*

…… 全金发放好骷髅，回到柯镇恶身边。六兄弟惘然望着**大哥**，静待他解说。只见他抬头向天，脸上肌肉不住扭动，森然道：“这是铜尸铁尸！”朱聪吓了一跳，道：“铜尸铁尸不早就死了吗，怎么还在人世？”……

…… Quan Jin-fa quickly placed the skull back in their original positions and returned to Ke Zhen’e. All of the Six Freaks’ eyes were upon **Big Brother** as they quietly waited for his explanation. “It’s Copper Corpse and Iron Corpse!” Ke Zhen’e looked as if he was looking up at the sky and his face twitched continuously. “But aren’t they dead? Can they still be alive?” This news shocked Zhu Cong tremendously. ……

Appendix C: Examples for Case Analysis in *White Deer Plain* in Section 5

In each example, units enclosed in brackets of the same shape refer to a same character, and units in brackets of different shapes refer to different characters. Because *White Deer Plain* has not been published in English, the English translations are for reference only. For brevity, we omit some extraneous spans by ellipses.

[Example 4-1]

(他)就把(Ø)怎样去请阴阳先生，怎么在雪地里撒尿，怎么发现那一坨无雪的慢坡地，怎么挖出怪物，以及拉屎伪造现场的过程详尽述说了一遍，然后问：“你听说过这号事没有？”姐夫朱先生静静地听完，眼里露出惊异的神光，不回答他的话，取来一张纸摊开在桌上，又把一支毛笔交给(嘉轩)说：“(你)画一画(你)见到的那个白色怪物的形状。”(嘉轩)捉着笔在墨盒里膏顺了笔尖，(Ø)有点笨拙却是十分认真地画起来，画了五片叶子，又画了秆儿把叶子连结起来，最终还是不无遗憾地憨笑着把笔交给姐夫：“我不会画画儿。”朱先生拎起纸来看着，像是揣摩一幅八卦图，忽然嘴一抿神秘地说：“小弟，你再看看(你)画的是什么？ ”(嘉轩)接过纸来重新审视一番，仍然憨憨地说：“基本上就是(我)挖出来的那个怪物的样子。”姐夫笑了，接过纸来对(嘉轩)说：“(你)画的是一只鹿啊！”(嘉轩)听了就惊诧得说不出话来，(Ø)越看(自己)刚才画下的笨拙的图画越像是一只白鹿。

(He) explained in detail (Ø) how to invited Yin Yang master, how to pee in the snow, how to find the irrigated farmlands without snow, how to dig out the monster, and how to shit and forge the scene, and then asked, “have you heard of this?” His brother-in-law Mr. Zhu listened to him quietly and showed a surprised look in his eyes. Without answering his words, Mr. Zhu took a piece of paper, spread it on the table, and handed a writing brush to (Jia-xuan) and said, “(you) draw the shape of the white monster (you) see.” (Jia-xuan) took the brush and smoothed the tip in the ink cartridge, and (Ø) drew somewhat clumsily but very seriously. He drew five leaves and drew a stalk to connect the leaves, and finally handed the brush to his brother-in-law with a regretful smile: “I don’t know how to draw.” Mr. Zhu picked up the paper and looked at it as if trying to figure out an eight-diagram pattern. Suddenly, he pursed his lips and said mysteriously, “little brother, you look at what (you) are drawing?” (Jia-xuan) took the paper and looked at it again, and still said fatuously, “it’s basically what the monster (I) dug out looks like.” His brother-in-law smiled, took the paper and said to (Jia-xuan), “(you) drew a deer!” (Jia-xuan) heard it and was too surprised to speak. The more (Ø) looked at the clumsy picture drawn by (himself) just now, the more it looked like a white deer.

[Example 4-2]

…… 斗争(三官庙老和尚)的大会第一次召开，……(年逾六旬的老和尚)被捆绑在戏楼后台的大柱子上，(他)万万没有料到(自己)会有如此劫数。(老和尚)把三官庙的几十亩土地租给附近村庄的农民，靠收取租粮过着神仙般的日子。(他)私订下一个规矩，每年夏秋两季交租要男人来，而秋末议定租地之事，却要女人来而不要男人。……

…… The meeting to fight against (the old monk in the Sanguan Temple) was held for the first time. …… (The sexagenarian old monk) was tied to a big pillar in the backstage of the theater. (He) never expected (himself) would suffer such a doom. (The old monk) rented dozens of mu of land in the Sanguan Temple to farmers in nearby villages and lived like an immortal by collecting the taxation cereal. (He) made a private rule that men are required to pay the rent in summer and autumn every year, but women are required to come instead of men when negotiating the renting land at the end of autumn. ……

[Example 4-3]

…… (黑娃)当即命令：“用炮轰！” …… (黑娃)当即驰马禀告团长，不料[一营长][白孝文]和{二营长}{焦振国}闻听炮声之后已赶到团部，立即报告了开炮的原因，而且极力鼓动团长调一营二营步兵去追击。…… <县党部书记><岳维山>亲自到会动员：……

…… (Little Black) immediately ordered: “bombard with artillery!” …… (Little Black) instantly galloped to report to the regimental commander. Unexpectedly, [the commander of the first battalion] [Bai Xiao-wen] and {the commander of the second battalion} {Jiao Zhen-guo} had rushed to the regimental headquarters after hearing the sound of artillery, immediately reported the reason for firing, and strongly encouraged the regimental commander to send infantry of the first and second battalions to pursue. …… <the Party branch secretary of the county> <Yue Wei-shan> personally attended the meeting to mobilize: ……

[Example 4-4]

(鹿三)扔下筷子，舀来一瓢凉水，让(嘉轩)漱口刷牙。…… 冷先生这时才得知(嘉轩)根本没有同母亲商量， …… (白赵氏)的心病不是那二亩水地能不能卖，而是这样重大的事情儿子居然敢于自作主张瞒着(她)就做了，自然是根本不把(她)当人了。(Ø)想到(秉德老汉)死没几年儿子就把(她)不当人，白赵氏简直都要气死了。

(Lu San) immediately put down his chopsticks and scooped a ladle of water to let (Jia-xuan) gargle. …… Only then did Mr. Leng know that (Jia-xuan) didn’t talk over with his mother before selling out farmlands. …… (Mrs. Bai) wasn’t angry for that the two mu irrigated farmlands were sold out, but for that his son should decide for himself without consulting (her). This behavior naturally indicated that his son didn’t treat (her) as one of their family. (Ø) Considering that it was only a few years after the death of (Old Man Bing-de), her son should ignore (her), Mrs. Bai almost went up the wall at the thought of this.

[Example 4-5]

(黑娃)问了人找着了(兆鹏)的房子。(兆鹏)穿着一条短裤正在擦洗身子，说：“啊呀稀客随便坐！”(兆鹏)出门泼了水回来蹬上长裤，给(黑娃)倒下一杯凉茶，俩人就聊起来。“(黑娃)(你)咋搞的？(Ø)也不来我这儿谝谝闲话？”“你忙着教书，(我)忙着打土坯挣钱，咱们都没闲空儿。”“(你)这两年日子过的咋样？”“(Ø)凑凑合合好着哩！”“(你)打短工挣的粮食够吃不够？”“(Ø)差不了多少够着哩！”“(你)住的那间窑洞浑全不浑全？”“(Ø)没啥大麻达倒塌不了！”“(你)百事如意哟！”(兆鹏)揶揄地说，随之刻意地问：

“(你)偷回来个媳妇族长不准(你)进祠堂拜祖，(你)心里受活不受活？(Ø)脸上光彩不光彩？” “(你)放屁！” (黑娃)像(Ø)遭到火烧水烫似的从椅子上弹起来，脸色骤变，“(你)当校长(Ø)闲烦了是不是？(Ø)想拿穷娃寻开心了是不是？”......

(Little Black) asked someone and found (Zhao-peng)’s house. (Zhao-peng) was wearing a pair of shorts and rubbing his body, and said, “Ah, a rare guest, just sit anywhere!” (Zhaopeng) went out to splash water back, put on his pants, and poured a cup of herbal tea for (Little Black), and the two started chatting. “What’s the matter with (you), (Black Baby)? Why don’t (Ø) come here to chat with me?” “You're busy teaching. (I)'m busy making money by ramming adobe. We both have no free time.” “How have (you) been these years? ” “(Ø) Not bad!” “Do (you) earn enough food by doing odd jobs?” “(Ø) It's almost enough!” “Is the cave (you) live in intact?” “(Ø) Nothing serious, it won’t collapse!” “Everything goes well with (you)!” (Zhao-peng) jokingly said, and then deliberately asked, “(You) committed adultery with a woman and take her back. The patriarch did not allow (you) to enter the ancestral hall to worship. Did (you) feel wrong? Did (Ø) lose face?” “(You) fart!” (Little Black) bounced up from the chair as if (Ø) had been burned and scalded, and his countenance suddenly changed. “Are (you) tired of (Ø) being the principal? Don't (Ø) want to have fun with a poor child?” ......