

Research and Analysis of Grammatical Error Correction Technology for Chinese Documents

Wei Jin¹, Feng Jiang², Xiulai Wang^{3*}, Ningling Ma³, Yutao Zhang³

¹Faculty of Computing, Harbin Institute of Technology, Harbin, China

²School of Future Technology, Nanjing University of Information Science & Technology, Nanjing, China

³Nanjing Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China

Email: *wangxiulai@126.com

How to cite this paper: Jin, W., Jiang, F., Wang, X., Ma, N. and Zhang, Y. (2024) Research and Analysis of Grammatical Error Correction Technology for Chinese Documents. *Journal of Computer and Communications*, 12, 202-223.
<https://doi.org/10.4236/jcc.2024.128013>

Received: June 8, 2024

Accepted: August 26, 2024

Published: August 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the widespread use of Chinese globally, the number of Chinese learners has been increasing, leading to various grammatical errors among beginners. Additionally, as domestic efforts to develop industrial information grow, electronic documents have also proliferated. When dealing with numerous electronic documents and texts written by Chinese beginners, manually written texts often contain hidden grammatical errors, posing a significant challenge to traditional manual proofreading. Correcting these grammatical errors is crucial to ensure fluency and readability. However, certain special types of text grammar or logical errors can have a huge impact, and manually proofreading a large number of texts individually is clearly impractical. Consequently, research on text error correction techniques has garnered significant attention in recent years. The advent and advancement of deep learning have paved the way for sequence-to-sequence learning methods to be extensively applied to the task of text error correction. This paper presents a comprehensive analysis of Chinese text grammar error correction technology, elaborates on its current research status, discusses existing problems, proposes preliminary solutions, and conducts experiments using judicial documents as an example. The aim is to provide a feasible research approach for Chinese text error correction technology.

Keywords

Chinese Text Error, Judicial Documents, Neural Network, Deep Learning, Transformer

1. Introduction

With the rapid development of the Internet and information technology, cou-

pled with the widespread use of the Chinese language globally, there has been a significant increase in both Chinese texts written by beginners and professional documents from various fields. This surge in written material has posed a significant challenge to traditional manual text proofreading methods. Chinese texts written by beginners of the Chinese language are more focused on correcting grammatical errors, whereas texts in special domains must conform to specific norms in terms of both formatting, content presentation, and grammatical specifications. Given the unique characteristics of the Chinese language, which belongs to the Sino-Tibetan language family, Chinese learners whose native language belongs to the Indo-European family often encounter a wide range of grammatical errors during their learning journey. In the professional realm, the construction of informationization has led to the production of numerous Chinese texts, which may contain deviations and grammatical errors due to various reasons. Quality control of these documents faces insurmountable challenges, often failing to achieve the desired standards due to several objective factors. An automated Chinese text grammar correction system can provide invaluable assistance to Chinese language learners, enhancing the efficiency of their learning process. Additionally, such a system can significantly reduce the labor and material resources required for manual proofreading in other domains.

With the advent of machine learning, deep learning, and artificial intelligence, various error detection and correction algorithms have emerged to automate the checking of critical information in documents. These algorithms can swiftly identify textual errors and aid in their rectification, making the research of related algorithms a focal point of interest. Chollampatt [1] and Zheng *et al.* [2] have respectively summarized their research on the types of errors encountered in Chinese and English Grammatical Error Correction (GEC) tasks. English GEC tasks typically target grammar, spelling, and collocation errors, while Chinese GEC tasks concentrate on four primary types of grammatical errors: redundancy, missing words, improper collocation, and incorrect word order. Rozovskaya and Dahlmeier [3] [4] developed an automated proofreading system for English texts using a classifier, achieving impressive classification results. Junczys-Dowmunt, Chollampatt, Rozovskaya, and others [5]-[7] introduced a novel approach by viewing the text proofreading task as a monolingual translation, converting incorrect text to correct text. They applied a phrase-based statistical machine translation (PBMT) method to text proofreading and grammatical error correction tasks. Nevertheless, these methods have limitations in efficiency. Currently, most grammatical error correction tasks are geared towards English, and Chinese texts present significant differences compared to English. Consequently, many existing methods are not suitable for Chinese text grammatical error correction tasks.

To enhance the writing of Chinese texts in specialized fields and assist Chinese language beginners in mastering its usage, it is imperative to undertake research on correcting grammatical errors in Chinese texts. This study begins by

clarifying the definition of textual syntax errors. Then, based on the fundamental principles of syntax error correction methods, it summarizes the current research status and applications of these methods. Furthermore, it analyzes the limitations of existing approaches. Finally, the study provides an outlook on the current situation and aims to offer practical research ideas for relevant scholars.

1.1. The Definition of Grammatical Error Correction Task

GEC is designed to automatically identify and rectify grammatical errors in text using computer algorithms. Ideally, a GEC system can efficiently analyze the context of the text, pinpoint and correct each grammatical error, and ultimately produce a grammatically correct text that preserves the original input text's semantics [8]. In cases where the input text is found to be grammatically flawless upon detection, it is directly outputted without any modifications.

Previous studies have emphasized that textual errors extend beyond mere grammatical mistakes [8] [9]. These errors are broadly classified into five levels: lexical, syntactic, semantic, chapter structure, and pragmatic. Among lexical errors, spelling mistakes are prevalent. For instance, "technology" might be mistakenly written as "technolgy" due to subjective factors. Subject-verb disagreement is another common issue, such as the misuse of "is" and "are". Syntactic errors often involve contextual inconsistencies, like using the wrong verb tense. For example, in the sentence "These two cats had a fight yesterday", "had" might be wrongly replaced with "have". Semantic errors typically arise from misspelling or inappropriate word usage, as in the sentence "Jack looked after his cat at pet hospital", where "after" is incorrectly substituted with "like". Chapter structure errors usually relate to the disorganization of the text. Discourse-level errors, on the other hand, are harder to pinpoint and correct because, while the text might be grammatically correct, its application might be inappropriate or violate cultural norms or values.

1.2. Grammatical Errors in Chinese Texts

As highlighted in the preceding section, much of the existing research on grammatical error correction has primarily focused on lexical, syntactic, and semantic errors, overlooking errors related to chapter structure and discourse level. Furthermore, a significant portion of these studies has been devoted to the correction of grammatical errors in English texts. When it comes to correcting Chinese grammatical errors, common mistake types include harmonic words, confusing sound words, reversing word order, missing words, and similar word errors.

However, besides Chinese texts authored by language beginners, there exists a vast array of Chinese texts spanning various specialized fields. For instance, in the judicial domain, documents like verdicts and rulings carry legal weight or ensure law enforcement. These documents exhibit a broader range of formats compared to ordinary texts, and their common syntax differs. From a structural perspective, judicial documents adhere to a relatively fixed format, boasting a stronger logic and structure than regular texts.

2. Literature Review

Currently, the mainstream of text research encompasses areas such as machine translation, question-and-answer dialogues. Within this landscape, the task of grammatical error correction remains a niche area. Initially, researchers tackled this challenge by developing a series of manual rule-based methods. However, with the advancement and widespread adoption of machine learning, there has been a noticeable shift in research trends. Statistical classifier-based and machine translation-based approaches have gradually emerged. In particular, the evolution of deep learning has propelled a transition in machine translation, moving from traditional statistical machine translation (SMT) methods to neural machine translation (NMT) techniques. Prominent among these are the RNN seq2seq model [10] [11], the attention mechanism [12] [13], the ConvS2S model [14], and the Transformer model [15], which have demonstrated impressive performance in related tasks.

To summarize, we categorize the initial manual rule-based strategies as the rule-based approach. The subsequent research methodologies are defined as the statistical classifier-based approach, the SMT-based approach, and the NMT-based approach, respectively.

2.1. The Rule-Based Approach

Initial research on semantic error correction primarily relied on artificial rules [16]-[19]. These rule-based approaches are highly effective for addressing specific error types that strictly adhere to syntactic rules [20], such as subject-verb disagreement, noun number errors (singular-plural), and tense misuse. This method offers several advantages:

- 1) The artificial rules are primarily crafted by experienced human linguists, ensuring accurate identification and correction of errors in the input text.
- 2) This approach obviates the need for additional corpus annotation, simplifying the correction process.
- 3) Artificial rules are highly interpretable, providing direct and valuable feedback to language learners.

This approach, which has been in use since the 1980s, has played a pivotal role in syntactic error correction research. Even today, numerous product-level systems utilize rule-based error correction methods.

However, this approach has its limitations. Primarily, the rules cover a finite number of error types, making it impossible to achieve comprehensive coverage of all error types present in a text. Especially for complex grammatical errors that involve contextual connections, such as various forms of word collocations, defining exhaustive correction rules is practically impossible. Furthermore, since linguistic experts are responsible for providing the error correction rules, the process is both labor-intensive and time-consuming.

2.2. The Statistical Classifier-Based Approach

Over the course of research in syntactic error correction tasks, corpus annota-

tions have been accumulated, facilitating the development of statistical classifier-based approaches [21]-[29]. Since article and preposition errors occur most frequently in English texts, initial research efforts focused on these areas. For these types of errors, the correction effectiveness of classifiers significantly surpasses that of traditional rule-based methods. In English, articles include “a”, “an”, and “the”. Incorrect article usage is considered a closed class error. For this, a predefined set is established, and the corresponding set of articles is {a, an, the}, referred to as the confusion set. After extracting common linguistic features, such as neighboring words, dependent syntax trees, lexical properties, etc., the text is transformed into a numerical representation. These features are then used as input for the classifier. Once trained, the classifier can detect, locate, and correct errors in the text. Given the varying validity of linguistic features across different word classes, researchers have proposed training separate classifiers for distinct grammatical errors [8]. Han *et al.* [22] selected a diverse large-scale corpus to train a classifier for detecting coronal errors, achieving impressive classification results with an accuracy of 88%. Tetreault and Chodorow [26] focused their research on prepositional errors and suggested using a maximum entropy model for error detection and correction.

However, the above methods typically rely solely on local contextual information and possess significant limitations. Specifically, these methods require that there is only one syntax error within a fixed window size of the text, thereby limiting their ability to correct multiple syntax errors. When multiple grammatical errors coexist within a text, it often becomes necessary to train multiple independent classifiers for different error types to ensure effective error correction. This approach can be problematic when dealing with texts written by language learners, which may contain various types of grammatical errors. Simply cascading these mutually independent classifiers for text correction may yield unsatisfactory results. Additionally, error types related to word collocation and word choice, due to the large number of target categories, can be challenging to define as a classification problem.

Furthermore, the classifier-based approach is not suitable for Chinese text error correction tasks, primarily because of the absence of word boundaries and the extensive character set in the Chinese language. Unlike English, which has numerous fixed phrases, Chinese words are short and not clearly spaced. This necessitates the consideration of contextual factors in the process of Chinese grammatical error correction, posing significant challenges for the classifier method. Given that the common character set in Chinese exceeds 5000 characters, establishing fixed error categories for a classifier-based approach to Chinese grammatical error correction would be a tedious and time-consuming process.

2.3. The SMT-Based Approach

A good syntactic error correction system ideally aims to locate and correct all syntactic issues without altering the original semantics of the input. Recently, machine translation-based methods have gradually become mainstream, super-

seding traditional approaches [30]-[42]. Brockett *et al.* [30] introduced a novel paradigm for error correction, grounded in monolingual translation tasks, and presented a phrase-based statistical machine translation (PBMT) model. Their method involved creating “error-correcting” parallel sentence pairs by deliberately introducing errors, such as incorrect usage of uncountable nouns, into grammatically correct texts. A machine translation model was then trained to correct these texts. Mizumoto *et al.* [43] employed a similar technique for Japanese grammatical error correction. They trained their model using parallel sentence pairs scraped from lang-8, a language learning social networking site. This dataset became a benchmark for grammatical error correction tasks. Additionally, the authors extended their research to English text error correction [44]. Their study revealed that varying sizes of training data have different impacts on correcting diverse types of grammatical errors. The results indicated that the PBMT model excels at addressing errors occurring within local contexts but performs less well with long-distance dependency information, struggling to effectively handle related errors. Yuan and Felice [32] presented a machine translation algorithm tailored for lexical decomposition, based on the CoNLL-2013 grammatical error correction task. Although a subsequent study found this algorithm to be marginally less effective than other classifiers for five specific types of textual errors, it nevertheless demonstrated its efficacy as a method for simultaneously addressing multiple interactive grammatical errors in machine translation tasks. Yuan *et al.* [35] merged a rule-based approach with a statistical machine translation method to devise a cascading error correction system. This system optimized error correction for the CoNLL-2014 grammatical error correction review task. Compared to CoNLL-2013, the CoNLL-2014 task tackled 28 distinct types of grammatical errors, further affirming machine translation as a versatile tool for grammatical error correction. Additionally, to enhance the system’s output of potential error correction results, the researchers integrated an external large-scale language model. This integration significantly boosted performance, aligning with findings from multiple studies that candidate reordering can enhance error correction tasks [45]-[48].

2.4. The NMT-Based Approach

In recent years, the steady progress of neural networks has led to remarkable achievements in natural language processing (NLP) tasks. These advancements have primarily targeted areas such as sentiment analysis, text implication, sequence annotation, language modeling, and automatic summarization. Notably, traditional machine translation techniques are gradually giving way to NMT methods [49], and this trend has extended to text grammatical error correction as well. Yuan *et al.* [35] were pioneers in adapting an NMT model for grammatical error correction. Their approach involved a codec-based RNN enhanced with an attention mechanism for superior feature extraction. Furthermore, they addressed the issue of unknown words (UNKs) by substituting them in the out-

put using an unsupervised word alignment model coupled with a word-level translation model. Xie *et al.* [36] introduced an innovative RNN codec model that operates at the character level. This model incorporates an N-gram language model within the decoder to calculate candidate prefix scores. This strategy effectively circumvents the challenge posed by unregistered words resulting from spelling errors.

Ji *et al.* [37] introduced a neural machine translation model featuring a nested attention mechanism. Experimental results indicate that this mechanism effectively integrates word-level and character-level feature information, leading to excellent performance in spelling error correction tasks. However, a significant challenge arises from the Exposure Bias problem, which affects both the model and those trained using maximum likelihood estimation methods. This bias can potentially lead to unsatisfactory error correction outcomes for the text. To address this issue, Sakaguchi *et al.* [39] abandoned conventional methods and adopted a reinforcement learning approach. They designed a model that uses the GLEU evaluation metric as a reward function. The results demonstrate that this approach enhances the model's error correction performance at the single-sentence level. This innovation illustrates the potential of reinforcement learning in improving NLP tasks, particularly in areas where traditional methods may be limited by factors such as Exposure Bias.

Convolutional Neural Networks (CNN) have indeed found widespread application in the field of grammatical error correction. Chollampatt *et al.* [1] put forward the argument that since most grammatical errors in text tend to occur within localized regions of a sentence, CNNs, with their strong ability to extract localized features, might prove more effective for error correction than RNNs. Additionally, by stacking multiple layers of convolution, Convolutional Neural Networks can also capture long-distance dependencies within the text. This makes CNNs a valuable tool in addressing the complexities of grammatical error correction.

2.5. Summary of the Current Situation

Rule-based methods are commonly used as an initial approach for error correction. While they offer advantages such as high accuracy, no need for corpus annotation, and good interpretability, they also have significant drawbacks. Formulating rules can be costly, and some errors may be irreparable using this method. Classifier-based error correction methods, which rely on classifiers and confusion sets, are dependent on the specific types of errors present in the text. This limitation is largely overcome by SMT-based methods. SMT methods require no additional linguistic input, automatically learn confusion sets from parallel data, and can target multiple complex error types with a single model. However, SMT methods face challenges, particularly the need for large-scale manually labeled parallel corpora, which contrasts with classifier-based methods that can operate directly on unlabeled corpora. Furthermore, studies have revealed that GEC systems often have poor generalization abilities, struggling to

access broader contexts such as source and target effectively. To address these limitations, researchers have applied RNN to automatic text proofreading and error correction, proposing a series of seq2seq models. These models have shown promise in improving GEC performance, particularly in handling complex error patterns and generalizing to new contexts.

Currently, the mainstream approach is to view grammatical error correction as a monolingual translation task, and the process of error correction is to “translate” the wrong sentence into the correct sentence. With the progress of neural machine translation, a large number of these methods and models have been migrated to the task of grammatical error correction, however, until 2018, the effect of the seq2seq series of models will not be able to surpass that of statistical-based machine translation models. Until Chollampatt *et al.* [1] proposed a multilayer-CNN-based model, the experimental results of this approach exceeded the neural network approach for statistical machine translation systems for the first time. In this paper, based on this, we use the more advanced model transformer to the Chinese grammar error correction task and get better results.

3. Methodology

Grammatical error correction is a typical task in NLP research. It aims to automatically locate and identify potential grammatical errors in text through algorithms and simultaneously correct them to produce accurate text.

Although the three methodological phases in related fields, as mentioned earlier, have limitations, the advent and progress of deep learning have presented opportunities for further refining these methods. To empower the model to encompass a wide array of contexts within the text, researchers have incorporated concepts like continuous space representation and nonlinear mapping from deep learning. In previous studies, RNNs have often been selected as the benchmark for developing novel machine translation algorithms and applying them to the task of correcting grammatical errors in text [35]-[37]. It has been discovered that RNNs, along with models like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), excel in capturing long-range dependencies in sequence modeling. Nevertheless, the task of text error correction necessitates computational decomposition of symbol positions within sequences, posing challenges for model parallelization.

3.1. Baseline Model for Grammatical Error Correction

The Transformer [15] is a codec model that relies on the Self-Attention mechanism. This mechanism guarantees interaction between any position in a sequence and other positions, enabling the Transformer to excel in capturing long-distance dependencies. Furthermore, the Transformer facilitates the parallelization of sequence computation. These strengths effectively overcome the limitations of RNNs and CNNs, explaining why, in the field of machine translation, Transformer-based models significantly outperform their RNN-based or

CNN-based counterparts [49]. In this study, we have chosen the Transformer as the foundation for designing our grammar error correction system. We will evaluate our system using a collected corpus of judicial documents and a publicly available Chinese text error correction corpus. Additionally, we will compare its performance to models based on RNN and CNN. We aim to formally define syntactic errors and detail specific methods for codec models utilizing RNNs and attention mechanisms in the context of syntactic error correction.

3.1.1. A Formal Definition of Grammatical Error Correction

In order to formally define the problem of sequence-to-sequence syntactic error correction, we describe the process in terms of a single “error-correcting” parallel sentence pair: the input text containing syntactic errors is corrected to correct text that does not contain syntactic errors and is output, while ensuring that the semantics of the original input text is not changed. Formally, given a source error sentence $x_i = \{x_1, x_2, \dots, x_m\}$ and a corresponding corrected target sentence $y_j = \{y_1, y_2, \dots, y_n\}$, each element in the set represents a word, and the syntax error correction modeling conditional probability distribution is expressed as follows:

$$p(y|x) = \prod_{t=1}^n p(y_t | x, y_{1:t-1}; \theta).$$

The model will generally learn the model parameters through great likelihood estimation, whose computational procedure can be described as:

$$\theta = \arg \max \sum_{t=1}^n \log p(y_t | x, y_{1:t-1}; \theta).$$

3.1.2. Recurrent Neural Network

RNN as a classical network model is often utilized for sequence modeling. Its fundamental principle lies in the ability to transform an input sequence into a vector that preserves the structural features of the original input. **Figure 1** illustrates a schematic representation of an RNN network, where the network unfolds progressively across time steps.

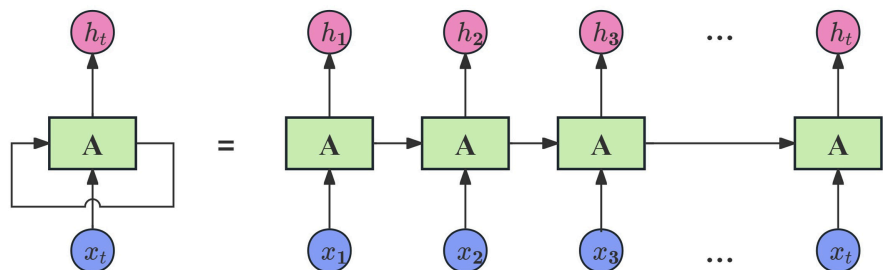


Figure 1. Recurrent neural networks expanded in time steps.

However, previous investigations have revealed that RNNs are susceptible to issues like gradient vanishing or gradient explosion during the training phase. This vulnerability arises from the network’s feed-forward computation, where each consecutive time step involves the application of a nonlinear activation

function. This process can cause the gradient to undergo significant fluctuations during back propagation. Gradient vanishing can hinder the capture of long-distance dependencies, while gradient explosion can result in gradient values escalating towards infinity. Factors such as an excessive number of network layers, an inappropriate loss function, or excessively large initial weights can contribute to these issues.

In order to solve the long-time dependency problem, researchers have proposed LSTM and GRU network models by deliberate design, and developed a series of variants in subsequent research. The core idea is to control the information flow in the network through a “gate” mechanism to selectively process the input sequences.

3.1.3. Encoder-Decoder Model and Attention Mechanisms

The Encoder-Decoder structure is a prevalent model framework in deep learning, and models employing this structure have proven highly successful in neural translation tasks. Additionally, in the field of computer vision, this model framework is widely used for tasks such as image segmentation and image captioning. Initially, codec models were primarily studied in the context of machine translation. Currently, state-of-the-art (SOTA) codecs often utilize LSTM networks as the backbone model, integrated into NMT-based systems. The Encoder-Decoder framework offers flexibility, accommodating various network models like CNN, LSTM, GRU, and others. This versatility extends to different modal inputs, including text, speech feature sequences, and images. Taking the NMT model as an example, the encoder converts the input text F into feature vectors through encoding. These feature vectors are then decoded by the decoder to produce the corrected text output. The structure of an RNN-based Encoder-Decoder model is depicted in **Figure 2**.

The input vector is $h_0^{(f)} = 0$, the model traverses the input text sequence and finally outputs the hidden state vector $h_{|F|}^{(f)}$. $h_{|F|}^{(f)}$ can be considered as the feature obtained by encoding all the information of the input text sequence. In the decoder, $RNN^{(e)}$ computes the feature vectors at each time step, which are computed by softmax function to get the required output for the text error correction task.

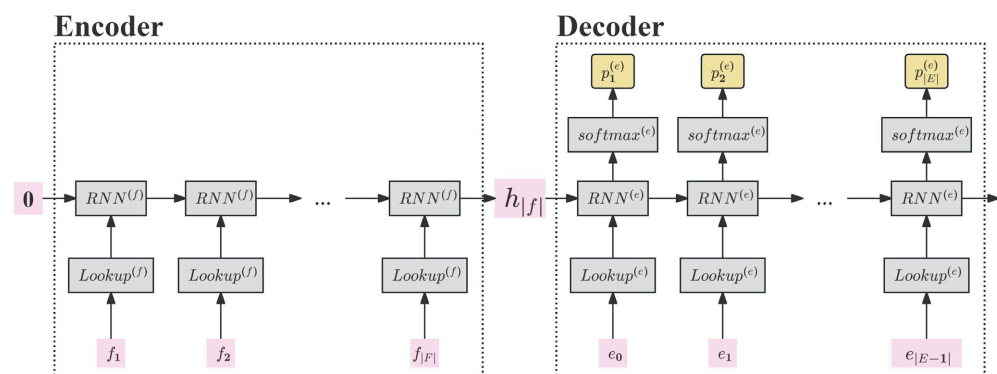


Figure 2. The structure of the RNN-based encoder-decoder model.

Attention mechanisms are inspired by the study of human visual and cognitive systems. In text-related tasks, this approach enables the extraction of necessary information from extensive text sequences. During the encoding process, features are preserved by encoding intermediate results, and the final output is determined by training the model to align with these preserved intermediate outputs. Unlike traditional codecs, the attention mechanism does not rely on a static representation of feature vectors; instead, it focuses on the pertinent information required for the current output. In a conventional encoder-decoder architecture, the encoder's output is either used as an initialization input for the decoder or provided as input at each step. Nevertheless, this structure faces limitations, such as the restricted number of encoder states and its inability to retain excessive information. Additionally, during the decoding process, the decoder's direct association with the incoming state at each step may not directly correlate with the network input. The introduction of the attention mechanism allows the encoder to adapt its focus to different inputs at various moments. The encoder-decoder model incorporating the attention mechanism is illustrated in **Figure 3**.

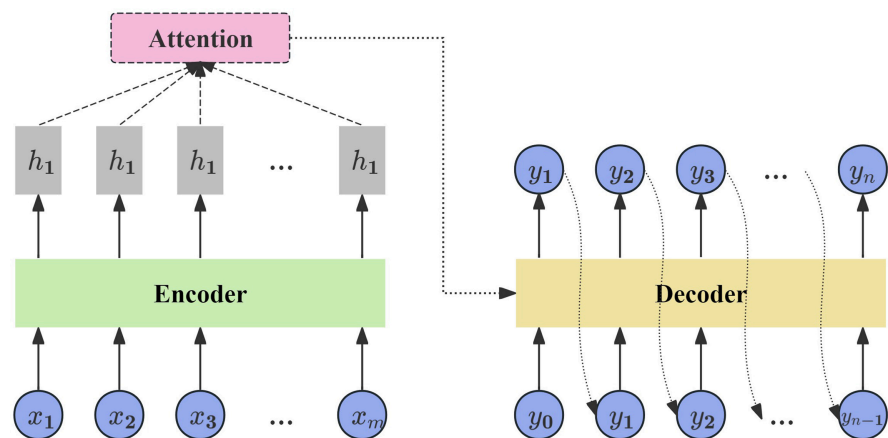


Figure 3. The encoder-decoder model with attention mechanism.

3.2. A Model for Correcting Grammatical Errors in Chinese Text Based on the Self-Attention Mechanism

It is shown that Transformer-based encoder-decoder model is significantly better than CNN-based and RNN-based encoder-decoder models on machine translation tasks. Therefore, this paper draws on the idea of converting a text grammar error correction task into a monolingual machine translation task for experimentation.

3.2.1. The Structure of Model

Figure 4 shows the model structure of Transformer, where the encoder and decoder are also present in the model. Define the error text at the source text as $x_i = \{x_1, x_2, \dots, x_m\}$. By encoding x_i into a set of implicit state representations $e_j = \{e_1, e_2, \dots, e_m\}$ in continuous space, the decoder generates the corrected

target sentence $y_k = \{y_1, y_2, \dots, y_n\}$ based on the implicit states time-step by time-step.

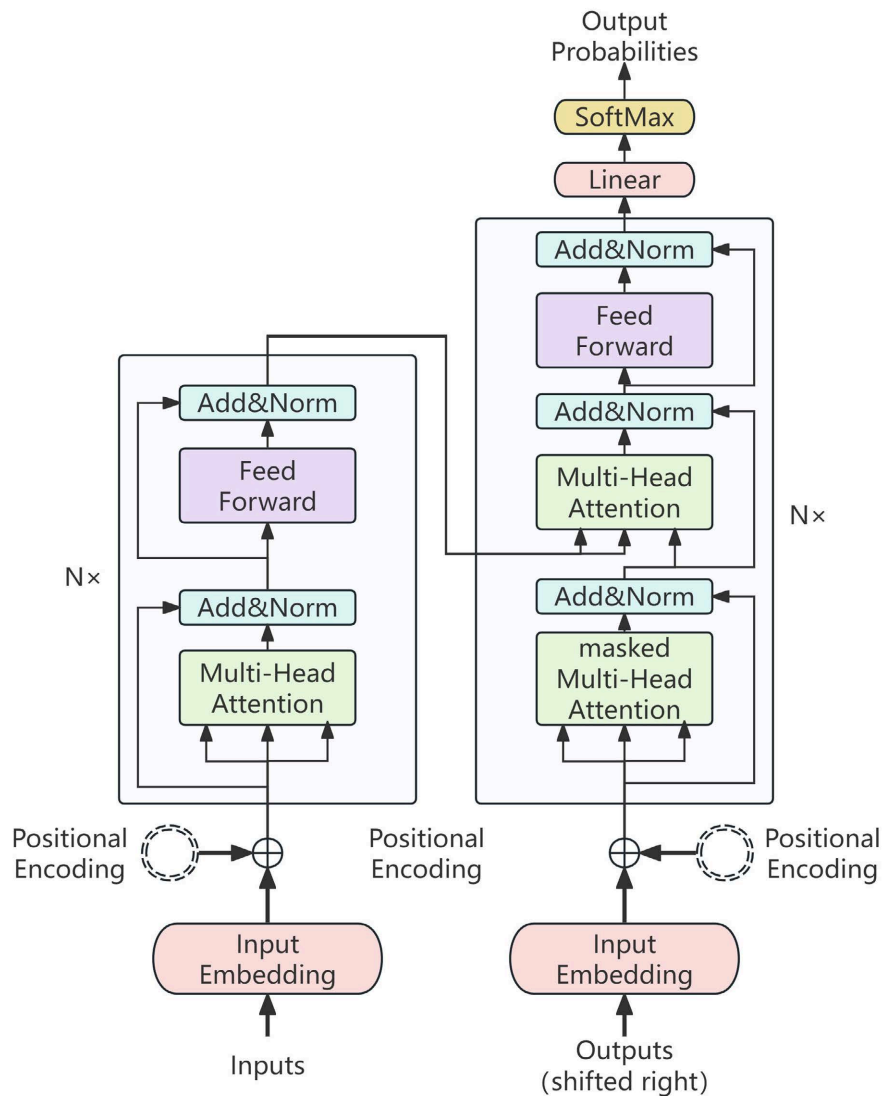


Figure 4. The structure of transformer.

The model designed in this paper consists of six identical layers of encoder and decoder. Each layer in the encoder consists of a self-attention module and a feed forward network module. The decoder follows a similar structure, including a masked self-attention module, a codec attention module, and a feed forward network module. Every module employs residual concatenation in its computations, followed by layer normalization.

During the encoding process, the self-attention mechanism within the encoder ensures optimal coding results by enabling specific positional symbols in the text sequence to interact with other positional symbols. Attention mechanisms are adept at capturing both long-distance and short-distance dependencies through the utilization of the query vector (Q), key vector (K), and value vector

(V). Multi-head self-attention, on the other hand, involves the use of different linear projections, obtained through independent learning, to compute Q, K, and V. These are then integrated, spliced, and transformed again via linear projections. Notably, Q, K, and V are all derived from the outputs of the previous layer.

To preserve the autoregressive property of the decoder, it is essential to mask information pertaining to the decoding position and beyond in the computational equation. It's important to note that Q originates from the output of the decoder's previous layer, while both K and V originate from the encoder's output.

Unlike RNN, there is no cyclic structure in the Transformer, and in order to ensure the effective utilization of the symbol position information in the sequence, the embedding often contains a position encoding, whose dimension is the same as the implied dimension of the model itself, which is computed by the formula as follows:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

where pos denotes the positional labeling of the symbol and i denotes a component of the positional encoding vector.

Typical sequence-to-sequence models, including the Transformer, often involve an embedding layer at the base of the encoder-decoder structure. This layer transforms input symbols into internal feature vectors. During the decoding process, to generate the target word, the model's output undergoes a linear transformation followed by a softmax function. This conversion produces a vector representing the probability distribution over the entire word list. The target word is then chosen based on a predefined threshold.

However, in the context of text syntax error correction tasks, the model described in this paper adopts a specific approach. Unlike some other models, it does not share parameters between the encoder, the embedding layer of the decoder, and the linear transformation layer preceding the softmax operation. This approach is tailored to the unique challenges of syntax error correction, where maintaining distinct parameters for these components can enhance the model's ability to accurately identify and correct errors in the text.

When training Transformer, great likelihood estimation is used with the goal of maximizing the likelihood of the model on the training data, whose computational procedure can be described as:

$$\theta = \arg \max \sum_{(x,y) \in S} \log p(y|x;\theta)$$

3.2.2. Decoding Strategy

For text x_i containing syntax errors, the Transformer model employs Beam Search decoding to generate corrected text y_{cor} . The top k_i candidates with the highest scores are eventually retained at each time step.

During the decoding process, the model may be biased towards outputting shorter texts, and in order to suppress this behavior, we introduce a length penalty term in the computation of the likelihood score, which is formulated as:

$$\text{score}(y_{cor}, x) = \frac{\log(p(y_{cor} | x))}{LP(y_{cor})},$$

$$LP(y_{cor}) = \frac{(5 + |y_{cor}|)^\alpha}{(5 + 1)^\alpha}, \alpha \in [0.6, 0.7].$$

3.3. Experimental Data

In conducting the experiments, we not only considered the error correction problem of ordinary Chinese texts, but also considered the grammatical error correction problem of special Chinese texts. Since judicial documents are representative of special Chinese texts, we chose them as samples of special Chinese texts to construct the dataset.

In this study, we utilized three primary training corpora. Two of these are extensive public datasets, while the third is a smaller, self-produced dataset focusing on judicial documents. For training purposes, we selected the HSK parallel corpus, the NLPCC 2018 GEC training set, and our specialized judicial document dataset. Additionally, to develop word vectors and statistical language models, we employed the Chinese Wikipedia corpus.

The NLPCC 2018 GEC training set originates from the Lang-8 multilingual learning platform. This platform offers a vast collection of articles authored by learners and subsequently revised by native speakers. From this repository, we retained 1,206,531 complete parallel sentence pairs for our training samples. This extensive dataset was featured in the NLPCC 2018 GEC shared task evaluation, making it the largest publicly accessible parallel corpus. An illustrative example from the NLPCC 2018 GEC database can be found in **Table 1**.

Table 1. The example of the NLPCC 2018 GEC database.

Incorrect Sentences in English	Sentences after Correction in English
Growing up, I had fun making lunch boxes.	I've had a lot of fun making boxed lunches as an adult.
People in the city can live a multifaceted life.	People in the city get to live colorful lives.
	People in the city are able to live a variety of lives.
	People in the city are able to live multifaceted lives.

The HSK corpus, fully named “HSK Dynamic Composition Corpus”, comprises essays penned by non-native Chinese speakers who took the HSK exam between 1992 and 2005. The initial version of this corpus encompassed 10,740 essays, totaling approximately 4 million words. Following modifications and additions, the corpus now includes 11,569 compositions, amounting to roughly 4.24 million words. It features 156,870 parallel sentence pairs, all of which have undergone thorough preprocessing, ensuring exceptional corpus quality.

Furthermore, utilizing large-scale natural language training samples as a foundation, we independently labeled 34,150 parallel sentence pairs from a small collection of judicial documents. These documents were sourced from publicly available online data. We segmented the dataset into training and testing sets using a 3:1 ratio.

Throughout our experiments, we experimented with various data fusion models for model training, yielding diverse experimental outcomes. **Table 2** illustrates the number of sentence pairs present in the different fused datasets we employed.

Table 2. The number of sentence pairs for different fused datasets.

Fused Dataset	The Number of Sentence Pairs
NLPCC + HSK	1,363,401
NLPCC + HSK + Judicial Documents	1,389,013
NLPCC + Judicial Documents	1,232,144
HSK + Judicial Documents	182,483

4. Research Results

4.1. Analysis of Experimental Results

To validate the performance of the models presented in this paper, we conducted an experimental comparison with state-of-the-art (SOTA) methods based on various models. To ensure a fair comparison, we excluded results from methods that rely on additional corpus resources. The models selected for comparison include MLConv, MLConv (4 ens), MLConv (4 ens) + EO, SMT + NNJM, and RNNSearch + align. MLConv is an encoder-decoder model that utilizes stacked multilayer convolution. MLConv (4 ens) is an ensemble model based on MLConv, while MLConv (4 ens) + EO is a system that further refines the candidate corrected texts from the syntactic error correction model using edit distance features. SMT + NNJM is a system model that adopts a statistical machine translation approach, combining a neural network model with a language model. RNNSearch + align is an RNN encoder-decoder model that incorporates an attentional mechanism. This approach addresses spelling errors in text by utilizing unsupervised word alignment models. We retrained these modeling approaches using our chosen experimental data and achieved experimental results comparable to those reported in the original studies. This retraining process allowed us to

directly compare the performance of our models with these SOTA methods under similar conditions.

We conducted tests on various fused datasets using the Transformer model. For professional documents, the optimal results were achieved with the fused dataset combining NLPCC, HSK, and Judicial Documents. These experiments demonstrate that the size of the dataset can indeed influence the effectiveness of the syntactic error correction method to a certain degree. Consequently, in our subsequent model comparison experiments, we selected the NLPCC + HSK + Judicial Documents fusion dataset as our training dataset. We evaluated its performance using both ordinary text and judicial test data as separate test sets.

The results of these evaluations are presented in **Table 3** and **Table 4**, respectively.

Table 3. The test results of different models (ordinary text data).

Method	Ordinary Text Data		
	Precision (%)	Recall (%)	F _{0.5} (%)
Transformer	60.42	34.01	52.30
MLConv	55.71	27.41	46.18
MLConv (4 ens)	63.14	25.62	48.84
MLConv (4 ens) + EO	59.31	31.42	50.37
SMT + NNJM	55.21	23.04	43.16
RNNSearch + align	53.13	21.42	40.99

Table 4. The test results of different models (judicial test data).

Method	Judicial Test Data		
	Precision (%)	Recall (%)	F _{0.5} (%)
Transformer	56.31	30.91	48.36
MLConv	53.62	24.16	43.11
MLConv (4 ens)	59.14	22.52	44.62
MLConv (4 ens) + EO	56.67	28.13	47.11
SMT + NNJM	52.13	20.68	39.97
RNNSearch + align	49.18	18.83	37.19

The experimental results demonstrate that the Transformer model significantly outperforms other methods. Compared to the Transformer, the equilibrium score of the MLConv model, which is also a single model, is approximately 5% lower. The ensemble model MLConv (4 ens) + EO exhibits the closest performance to the Transformer. Furthermore, upon evaluating the metrics, it becomes evident that the performance of RNN-based models is not as good as that of multilayer CNN models. This is primarily because most grammatical errors in text occur locally, and CNNs possess a stronger capability to capture local fea-

tures when compared to RNNs.

The experiments also reveal that even after incorporating some judicial document text data into the training set, the model performance metrics experience a slight degradation.

In summary, treating Chinese text syntax error correction as a machine translation problem and employing the Transformer model is a feasible approach. Indeed, this method can surpass existing syntax error correction techniques in terms of effectiveness. Additionally, this approach circumvents the need for a vast number of manually crafted rules, which is a requirement in rule-based grammatical error correction methods. However, the issue of corpus size within specific domains remains a significant factor limiting the effectiveness of this type of approach.

4.2. Discussion

Since the introduction of sequence-to-sequence learning in 2014 and its subsequent success in machine translation, this learning paradigm has been extensively adopted for various generative tasks, including summarization and question-answering. However, one limitation of this approach has been the encoder's representation of the input sequence as a fixed-size vector, which often results in the inefficient utilization of contextual information, thereby compromising the effectiveness of model learning. The advent of the attention mechanism revolutionized this by enabling the model to selectively attend to different parts of the input sequence, thereby capturing and utilizing contextual information more effectively. Due to the significant performance enhancements achieved through this mechanism, integrating attention into sequence-to-sequence learning has become a standard practice in natural language generation tasks. Despite the Transformer model's superior performance compared to other seq2seq models, making it a prevalent method in machine translation research, experimental results indicate that its accuracy, when applied solely to syntax error correction systems, is not yet sufficient. This suggests that further improvements and innovations are necessary to enhance the model's effectiveness in this specific task.

Currently, pre-training modeling techniques are extensively employed across various research domains. The experimental approach utilized in this paper relies on generating pre-training models from a vast amount of standard natural language text. By integrating a relatively smaller quantity of judicial data for further training, we produce the final model results. The achieved outcomes exhibit significant improvement compared to models trained solely on judicial document corpora. This methodology demonstrates the effectiveness of leveraging large-scale general language data for pre-training, followed by fine-tuning with specialized data, to enhance performance in domain-specific tasks where annotated resources are limited.

During our experiments, we discovered that certain segmentation tools improperly segment the fixed French phrases in judicial documents, leading to ambiguity. Finding a tool that accurately handles subjunctive terminology could

significantly enhance the effectiveness of our grammatical error correction model for professional documents within specialized domains. Given the abundance of specialized terms, exploring a disambiguation approach distinct from that used in standard natural language texts may improve the efficacy of our grammatical error correction method for specialized instruments. However, this would undoubtedly require extensive research into the semantic aspects of the terminology, which is a challenging but potentially rewarding task.

5. Conclusions

Integrating NLP research advancements in machine learning and deep learning into Chinese text syntax error correction methods is an inevitable trend, aiming to enhance efficiency and accuracy. While rule-based approaches are currently prevalent for addressing grammatical errors in Chinese text, they pose challenges in the professional domain. As the field evolves, the rules must adapt, requiring constant updates. Moreover, formulating a large number of rules demands considerable human and material resources, rendering this approach unsustainable in the long run.

In this study, we selected Transformer as the model architecture for the Chinese text error correction task. We first constructed a large dataset containing the NLPCC 2018 GEC dataset and HSK dataset, as well as judicial document data. We selected different data fusion patterns to fine-tune the pre-trained model and found that the data size can affect the text error correction ability of the model. Further experiments using judicial instruments revealed that specialized instruments may lead to degradation of model performance, limiting the performance gains of such methods. We believe that the impact of this problem can be mitigated through further data enhancements.

Although the SMT-based methods exhibit superior error correction performance compared to statistical learning classifiers, they still encounter several challenges. Firstly, numerous previous studies have revealed issues with the computation of maximum likelihood estimation. The calculation of the loss function during the training phase relies on prediction performance at the word level. However, due to the nature of text error correction tasks, evaluation metrics focused on prediction performance at the phrase or sentence level are commonly employed during testing.

Secondly, exposure bias is a prevalent issue in sequence-to-sequence learning models. During the prediction phase, the output of each time step serves as the input for the next time step. If an error occurs in one time step, it can cascade into subsequent time steps, leading to a series of incorrect predictions that are difficult to rectify. This phenomenon underscores the fragility of such models and the need for further improvements to enhance their robustness and accuracy.

It is undeniable that the practical application of this technical research is still a way off. The computation of extreme likelihood estimates may result in a disconnect between the training and prediction phases. In this context, the concept

of generative adversarial networks (GANs) could be applied to the task of textual error correction. A neural grammar error correction model can be considered to have excellent error correction performance when its corrected text output is indistinguishable from human-written corrected text. This approach offers a promising direction for bridging the gap between machine-generated and human-like text corrections, potentially leading to more natural and accurate outputs in the field of textual error correction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Chollampatt, S. and Ng, H.T. (2018) A Multilayer Convolutional Encoder-Decoder Neural Network for Grammatical Error Correction. *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 5755-5767. <https://doi.org/10.1609/aaai.v32i1.12069>
- [2] Zheng, B., Che, W., Guo, J., et al. (2016) Chinese Grammatical Error Diagnosis with Long Short-Term Memory Networks. *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2016)*, Osaka, 12 December 2016, 49-56.
- [3] Rozovskaya, A., Chang, K., Sammons, M., Roth, D. and Habash, N. (2014) The Illinois-Columbia System in the CoNLL-2014 Shared Task. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, Baltimore, 26-27 July 2014, 34-42. <https://doi.org/10.3115/v1/w14-1704>
- [4] Dahlmeier, D., Ng, H.T. and Ng, E.J.F. (2012) NUS at the HOO 2012 Shared Task. *Proceedings of the 7th Workshop on Building Educational Applications Using NLP*, Montréal, 3-8 June 2012, 216-224.
- [5] Junczys-Dowmunt, M. and Grundkiewicz, R. (2016) Phrase-Based Machine Translation Is State-of-the-Art for Automatic Grammatical Error Correction. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, 1-5 November 2016, 1546-1556. <https://doi.org/10.18653/v1/d16-1161>
- [6] Chollampatt, S. and Ng, H.T. (2017) Connecting the Dots: Towards Human-Level Grammatical Error Correction. *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, Copenhagen, 8 September 2017, 327-333. <https://doi.org/10.18653/v1/w17-5037>
- [7] Rozovskaya, A. and Roth, D. (2016) Grammatical Error Correction: Machine Translation and Classifiers. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, 7-12 August 2016, 2205-2215. <https://doi.org/10.18653/v1/p16-1208>
- [8] Yuan, Z. (2017) Grammatical Error Correction in Non-Native English. University of Cambridge.
- [9] Kukich, K. (1992) Techniques for Automatically Correcting Words in Text. *ACM Computing Surveys*, **24**, 377-439. <https://doi.org/10.1145/146370.146380>
- [10] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014,

- 1724-1734. <https://doi.org/10.3115/v1/d14-1179>
- [11] Sutskever, I., Vinyals, O. and Le, Q.V. (2014) Sequence to Sequence Learning with Neural Networks. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, 8-13 December 2014, 3104-3112.
- [12] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. arXiv: 1409.0473. <https://doi.org/10.48550/arXiv.1409.0473>
- [13] Luong, T., Pham, H. and Manning, C.D. (2015) Effective Approaches to Attention-Based Neural Machine Translation. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 17-21 September 2015, 1412-1421. <https://doi.org/10.18653/v1/d15-1166>
- [14] Gehring, J., Auli, M., Grangier, D., et al. (2017) Convolutional Sequence to Sequence Learning. *Proceedings of the 34th International Conference on Machine Learning*, Sydney, 6-11 August 2017, 1243-1252.
- [15] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [16] Macdonald, N., Frase, L., Gingrich, P. and Keenan, S. (1982) The Writer's Workbench: Computer Aids for Text Analysis. *IEEE Transactions on Communications*, **30**, 105-110. <https://doi.org/10.1109/tcom.1982.1095380>
- [17] Bustamante, F.R. and León, F.S. (1996) GramCheck: A Grammar and Style Checker. *Proceedings of the 16th Conference on Computational Linguistics*, Copenhagen, 5-9 August 1996, 175-181. <https://doi.org/10.3115/992628.992661>
- [18] Heidorn, G.E., Jensen, K., Miller, L.A., Byrd, R.J. and Chodorow, M.S. (1982) The EPISTLE Text-Critiquing System. *IBM Systems Journal*, **21**, 305-326. <https://doi.org/10.1147/sj.213.0305>
- [19] Richardson, S.D. and Braden-Harder, L.C. (1988) The Experience of Developing a Large-Scale Natural Language Text Processing System: CRITIQUE. *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, 9-12 February 1988, 195-202. <https://doi.org/10.3115/974235.974271>
- [20] Sakaguchi, K. (2018) Robust Text Correction for Grammar and Fluency. Johns Hopkins University.
- [21] Knight, K. and Chander, I. (1994) Automated Postediting of Documents. *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, Seattle, 1-4 August 1994, 779-784.
- [22] Han, N.R., Chodorow, M. and Leacock, C. (2004) Detecting Errors in English Article Usage with a Maximum Entropy Classifier Trained on a Large, Diverse Corpus. *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, 26-28 May 2004, 1625-1628.
- [23] Chodorow, M., Tetreault, J.R. and Han, N. (2007) Detection of Grammatical Errors Involving Prepositions. *Proceedings of the Fourth ACL-SIGSEM Workshop on Prepositions*, Prague, 28 June 2007, 25-30. <https://doi.org/10.3115/1654629.1654635>
- [24] De Felice, R. and Pulman, S.G. (2007) Automatically Acquiring Models of Preposition Use. *Proceedings of the Fourth ACL-SIGSEM Workshop on Prepositions*, Prague, 28 June 2007, 45-50. <https://doi.org/10.3115/1654629.1654639>
- [25] De Felice, R. and Pulman, S.G. (2008) A Classifier-Based Approach to Preposition and Determiner Error Correction in L2 English. *Proceedings of the 22nd International Conference on Computational Linguistics*, Manchester, 18-22 August 2008, 169-176. <https://doi.org/10.3115/1599081.1599103>

- [26] Tetreault, J.R. and Chodorow, M. (2008) The Ups and Downs of Preposition Error Detection in ESL Writing. *Proceedings of the 22nd International Conference on Computational Linguistics*, Manchester, 18-22 August 2008, 865-872. <https://doi.org/10.3115/1599081.1599190>
- [27] Han, N.-R., Tetreault, J., Lee, S.-H., et al. (2010) Using an Error-Annotated Learner Corpus to Develop an ESL/EFL Error Correction System. *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Valletta, 17-23 May 2010, 763-770.
- [28] Leacock, C., Chodorow, M., Gamon, M., et al. (2010) Automated Grammatical Error Detection for Language Learners. Morgan & Claypool.
- [29] Tetreault, J., Foster, J. and Chodorow, M. (2010) Using Parse Features for Preposition Selection and Error Detection. *Proceedings of the ACL 2010 Conference Short Papers*, Uppsala, 11-16 July 2010, 353-358.
- [30] Brockett, C., Dolan, W.B. and Gamon, M. (2006) Correcting ESL Errors Using Phrasal SMT Techniques. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the ACL*, Sydney, 17-18 July 2006, 249-256. <https://doi.org/10.3115/1220175.1220207>
- [31] Dahlmeier, D. and Ng, H.T. (2012) A Beam-Search Decoder for Grammatical Error or Rection. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Jeju Island, 12-14 July 2012, 568-578.
- [32] Yuan, Z. and Felice, M. (2013) Constrained Grammatical Error Correction Using Statistical Machine Translation. *Proceedings of the Seventeenth Conference on Computational Natural Language Learning. Shared Task*, Sofia, 8-9 August 2013, 52-61.
- [33] Chollampatt, S., Hoang, D.T. and Ng, H.T. (2016) Adapting Grammatical Error Correction Based on the Native Language of Writers with Neural Network Joint Models. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, 1-5 November 2016, 1901-1911. <https://doi.org/10.18653/v1/d16-1195>
- [34] Chollampatt, S., Taghipour, K. and Ng, H.T. (2016) Neural Network Translation Models for Grammatical Error Correction. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, New York, 9-15 July 2016, 2768-2774.
- [35] Yuan, Z. and Briscoe, T. (2016) Grammatical Error Correction Using Neural Machine Translation. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies*, San Diego, 12-17 June 2016, 380-386. <https://doi.org/10.18653/v1/n16-1042>
- [36] Xie, Z., Avati, A., Arivazhagan, N., et al. (2016) Neural Language Correction with Character-Based Attention. arXiv: 1603.09727. <https://doi.org/10.48550/arXiv.1603.09727>
- [37] Ji, J., Wang, Q., Toutanova, K., Gong, Y., Truong, S. and Gao, J. (2017) A Nested Attention Neural Hybrid Model for Grammatical Error Correction. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, 30 July-4 August 2017, 753-762. <https://doi.org/10.18653/v1/p17-1070>
- [38] Napoles, C. and Callison-Burch, C. (2017) Systematically Adapting Machine Translation for Grammatical Error Correction. *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, Copenhagen, 8 Sep-

- tember 2017, 345-356. <https://doi.org/10.18653/v1/w17-5039>
- [39] Sakaguchi, K., Post, M. and Van Durme, B. (2017) Grammatical Error Correction with Neural Rein-Force ment Learning. *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Taipei, 27 November-1 December 366-372.
- [40] Junczys-Dowmunt, M., Grundkiewicz, R., Guha, S. and Heafield, K. (2018) Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, 1-6 June 2018, 595-606. <https://doi.org/10.18653/v1/n18-1055>
- [41] Grundkiewicz, R. and Junczys-Dowmunt, M. (2018) Near Human-Level Performance in Grammatical Error Correction with Hybrid Machine Translation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, New Orleans, 1-6 June 2018, 284-290. <https://doi.org/10.18653/v1/n18-2046>
- [42] Zhao, W., Wang, L., Shen, K., Jia, R. and Liu, J. (2019) Improving Grammatical Error Correction via Pre-Training a Copy-Augmented Architecture with Unlabeled Data. *Proceedings of the 2019 Conference of the North*, Minneapolis, 2-7 June 2019, 156-165. <https://doi.org/10.18653/v1/n19-1014>
- [43] Mizumoto, T., Komachi, M., Nagata, M., et al. (2011) Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. *Proceedings of 5th International Joint Conference on Natural Language Processing*, Chiang Mai, 8-13 November 2011, 147-155.
- [44] Mizumoto, T., Hayashibe, Y., Komachi, M., et al. (2012) The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings. *Proceedings of COLING 2012: Posters*, Mumbai, 8-15 December 2012, 863-872.
- [45] Junczys-Dowmunt, M. and Grundkiewicz, R. (2014) The AMU System in the Conll-2014 Shared Task: Grammatical Error Correction by Data-Intensive and Feature-Rich Statistical Machine Translation. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, Baltimore, 26-27 July 2014, 25-33. <https://doi.org/10.3115/v1/w14-1703>
- [46] Mizumoto, T. and Matsumoto, Y. (2016) Discriminative Reranking for Grammatical Error Correction with Statistical Machine Translation. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, 12-17 June 2016, 1133-1138. <https://doi.org/10.18653/v1/n16-1133>
- [47] Yuan, Z., Briscoe, T. and Felice, M. (2016) Candidate Re-Ranking for SMT-Based Grammatical Error Correction. *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, San Diego, 12-17 June 2016, 256-266. <https://doi.org/10.18653/v1/w16-0530>
- [48] Hoang, D.T., Chollampatt, S. and Ng, H.T. (2016) Exploiting N-Best Hypotheses to Improve an SMT Approach to Grammatical Error Correction. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, New York, 9-15 July 2016, 2803-2809.
- [49] Wu, Y., Schuster, M., Chen, Z., et al. (2016) Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv: 1609.08144. <https://doi.org/10.48550/arXiv.1609.08144>