



## An Automatic Correction Model for Subject Verb Agreement Errors in Second Language Learners Based on Deep Learning

Yueyi Liu<sup>1,\*</sup>

<sup>1</sup> Nanfang College, Guangzhou, Guangzhou 510900, Guangdong, China

**SUMMARY:** *This paper researches a relatively common problem among learners of second languages mistakes of subject and verb agreement. Human beings have attempted to handle this problem in previous times, generally through the utilization of manually made rules or statistical models. These methods may have okay effectiveness for relatively simple sentences, but once matters become a little more complex or the environment changes, they therefore cannot always maintain very good effects. In order to circumvent that problem, we have constructed an alternative type of framework which introduces several concepts simultaneously instead of adhering to only a single one. It is composed of three component parts. There exists a component which is based on constraint, it marks the positions where the agreement could possibly have problems. there exists a kind of agent which travels inside the sentence structure, therefore it endeavors to comprehend how every part combines together in the syntactic aspect. In addition to this, a probability based filtering step examines possible revisions and evaluates which of these appear comparatively more probable. This system does not completely rely on rules, nor does it purely rely on data, therefore it mixes these two aspects with a certain method. In the actual operation, this therefore makes it a little more stable in different situations. When we carried out some experiments, the obtained outcomes were obviously superior to that which earlier methods have reached correct rate increased by over 15% on standard benchmark data collections. Certainly, this is not a perfect solution, but it does give indication that the bringing together of structural restrictions and probabilistic inference can provide help. At the very least up to current moment, this can be regarded as a feasible method that can promote the promotion of grammatical error correction work, particularly for language learners. It also can possibly provide certain guidance for the future research about automatic language studying tools.*

**KEYWORDS:** *business English writing, automated writing assessment, personalized feedback, discourse analysis, writing quality evaluation*

## 1 Introduction

Subject verb agreement mistakes show up pretty often when people are learning a second language [1]. It's one of those things that doesn't always stop you from understanding the sentence, but it can still make the writing feel a bit off [2]. Sometimes you have to pause for a second to figure out what's going on, and that alone can affect how smoothly the text reads [3]. fairly or not, make the writer seem less confident in the language. as more communication happens across different languages, there's been more attention on tools that can help with this kind of problem. Automated correction tools, for example, are starting to

\*liuyy7955@163.com

<https://doi.org/10.65102/is2026536>

play a bigger role[4].

Capturing writing errors is quite practical, this content is related to natural language processing work, and is especially applicable to system grammar error handling [5]. Early attempts to correct subject-verb agreement errors used the method of clarifying grammatical rules [6]. Rule-based systems are relatively simple and easy to understand, because language knowledge is integrated into the correction [7]. But the effect of practical application is often limited, especially when encountering irregular situations or exceptions of natural language [8]. Another problem is that the cost of maintaining such systems is relatively high. Designing or updating rules requires a lot of manual operations, making these methods difficult to apply to broader language contexts or different learner groups[9].

Later research explored statistical methods that use learning consistent patterns from labeled data to improve flexibility[10-12]. Methods like decision trees and support vector machines (SVMs) [13] can capture patterns of language use without complete manual rules [14]. These methods adapt better to unseen data than earlier systems. However, statistical models still have difficulty capturing complex syntactic dependencies, and performance often relies on handcrafted features, which limits the modeling of the full context of sentences [15]. Recently, deep learning methods have been widely used in grammatical error correction [16]. Neural models, especially those based on pre-trained language representations such as BERT and GPT, can learn complex language patterns from a large amount of text data [17]. This saves the need to manually construct features, and often can improve performance. However, these models also create some new challenges [18]. They usually require a large amount of training data and a large amount of computing resources, which may limit their application in places with insufficient resources or scarce data [19, 20].

Motivated by these limitations, this paper proposes a framework that combines the strengths of deep learning with structured linguistic guidance. The goal is to improve correction accuracy while reducing reliance on large scale resources. In particular, the proposed approach integrates neural architectures with efficient training strategies, allowing the model to better handle diverse sentence structures without significantly increasing computational cost. This design makes the system more practical for real world applications, especially in educational settings where resources may be limited. The main contributions of this work can be summarized as follows:

- We put together a framework that mixes neural modeling with some more explicit linguistic constraints. The idea was mainly to make it deal a bit better with those tricky agreement patterns, which don't always behave nicely.
- Use a proxy mechanism to dynamically explore syntactic structures. It cannot solve all problems, but it can help the model adjust when the sentence changes.
- Experiments show that this method is generally better than existing methods, the gain is not only reflected in accuracy, its scalability also has stable improvement, which is quite encouraging.

## 2 Method

### 2.1 Overview

This section explains how the model handles subject verb agreement errors in learner writing. Although the approach relies on deep learning, it does not treat correction as a single end to end task. Instead, the model separates three operations: detecting possible agreement errors, identifying the syntactic relations that matter for agreement, and selecting the best correction

in context. We adopted this design because a single module tended to learn frequent local patterns but was less reliable on sentences with longer dependencies or less common structures. Dividing these functions made the model easier to control and reduced training instability. Section 2.2 defines the task from both linguistic and computational perspectives. It identifies the syntactic environments in which agreement must be resolved and the learner errors the model is designed to correct. These include cases with intervening nouns, complex noun phrases, and other structures in which nearby words can distract the model from the true subject. Section 2.3 presents the Agreement Optimization Planner, which contains three modules: the Constraint Driven Error Corrector, the Agent Based Syntax Navigator, and the Probabilistic Agreement Filter. The Constraint Driven Error Corrector applies explicit agreement constraints to detect likely violations. The Agent Based Syntax Navigator analyzes sentence structure to determine which noun phrase controls the verb. The Probabilistic Agreement Filter then compares candidate corrections and ranks them by contextual plausibility.

## 2.2 Preliminaries

This section formalizes the subject verb agreement problem in second language learner texts and provides the necessary background for the proposed approach. The task focuses on identifying and correcting cases where the subject and verb do not agree in number, which remains a common issue due to the complexity of syntactic structures and the variability of learner errors. Let  $S = \{s_1, s_2, \dots, s_n\}$  denote a set of sentences produced by second language learners. Each sentence  $s_i$  is represented as a sequence of tokens  $w_{i1}, w_{i2}, \dots, w_{im}$ . The goal is to transform each  $s_i$  into a corrected sentence  $s'_i$  such that the agreement between the subject and verb is satisfied. To describe this process, we define a mapping function  $f: S \rightarrow S'$ , where  $S' = \{s'_1, s'_2, \dots, s'_n\}$  is the set of corrected sentences. The objective of  $f$  is to reduce the number of agreement errors across all sentences. We introduce an indicator function  $\delta(s_i)$ , which takes the value 1 if a sentence contains an agreement error and 0 otherwise. The overall objective can then be written as:

$$\min \sum_{i=1}^n \delta(s'_i)$$

To ensure that corrections are valid, it is necessary to identify the subject and verb in each sentence. Let  $\text{Subj}(s_i)$  and  $\text{Verb}(s_i)$  denote the subject and verb of sentence  $s_i$ , respectively. The agreement condition is defined as:

$$\text{Agree}(\text{Subj}(s_i), \text{Verb}(s_i)) = \begin{cases} 1, & \text{if the subject and verb agree in number} \\ 0, & \text{otherwise} \end{cases}$$

The corrected sentences are expected to satisfy the following condition:

$$\text{Agree}(\text{Subj}(s'_i), \text{Verb}(s'_i)) = 1, \quad \forall i$$

To model syntactic relationships, we rely on dependency parse trees, denoted as  $T(s_i)$  for each sentence. This makes it easier to track how tokens are connected grammatically, instead of leaving everything to the encoder. we consider dependency links of the form  $\text{Dep}(w_{ij}, w_{ik})$ , which describe how two tokens are related. Among these, the subject verb relation is the main focus here, so we extract it explicitly as follows:

$$\text{Dep}(\text{Subj}(s_i), \text{Verb}(s_i)) = \text{nsubj}$$

We refer to the overall framework as the Agreement Optimization Planner. It is not a single monolithic model, but rather a combination of three parts: the Constraint Driven Error Corrector, the Agent Based Syntax Navigator, and the Probabilistic Agreement Filter. The first applies linguistic constraints to detect possible errors, the second explores syntactic structures to generate candidate corrections, and the third evaluates these candidates using probabilistic criteria. The correction task can be formulated as a constrained optimization problem, where the goal is to minimize agreement errors while ensuring that grammatical constraints are satisfied. This can be expressed as:

$$\min \sum_{i=1}^n \delta(s'_i) \quad \text{subject to} \quad \text{Agree}(\text{Subj}(s'_i), \text{Verb}(s'_i)) = 1$$

The constraints are derived from syntactic rules and dependency structures. The optimization process is guided by an agent based exploration strategy, which iteratively refines candidate corrections based on feedback from the probabilistic filtering module.

### 2.3 Agreement Optimization Planner

As illustrated in Figure 1, this subsection describes the Agreement Optimization Planner, which is designed to correct subject verb agreement errors using a combination of structural constraints and data driven methods. The model is organized around three interconnected modules, each responsible for a different aspect of the correction process. These components together support more accurate identification and refinement of grammatical errors.

#### Methodology to Address Subject-Verb Agreement Errors

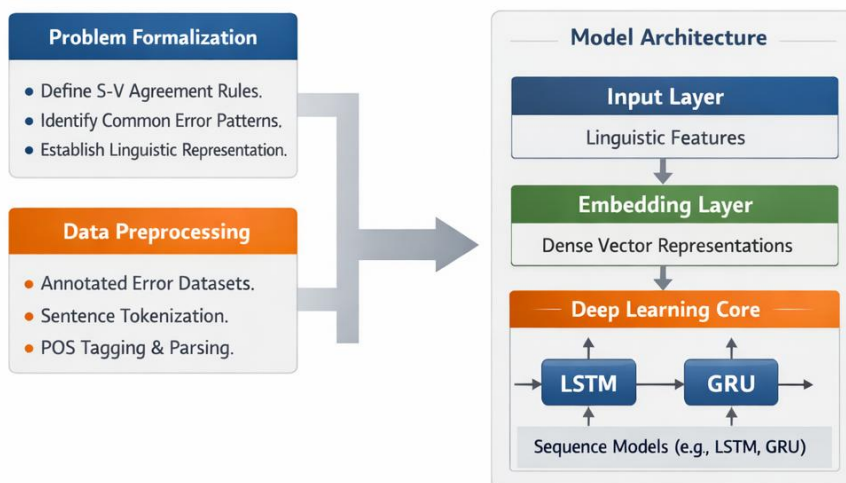


Figure 1: This method looks at subject verb agreement errors in L2 learners. It spells out what counts as agreement and points to typical issues like missing third person -s or attraction errors. Sentences are described using basic linguistic features.

**Constraint Based Error Correction Framework.** Figure 2 outlines the agreement correction step. We do not predict verb forms directly. For each sentence we enumerate a small set of candidates and score them with a set of constraints.

Let  $S = \{s_1, \dots, s_n\}$ , where each sentence is

$$s_i = \{w_{i1}, \dots, w_{im}\}$$

We extract a subject  $u_i$  from  $s_i$ :

$$u_i = \phi(s_i)$$

Candidate verb forms are collected into  $V_i \subseteq V$ . In practice,  $|V_i|$  is small usually 2 or 3, and rarely exceeds 4 even for irregular verbs.

We define a constraint set  $C = \{c_1, \dots, c_k\}$ . Each constraint acts as a binary check:

$$c_j(u_i, v) \in \{0, 1\}$$

with the corresponding indicator

$$\delta(c_j, u_i, v) = \begin{cases} 1, & c_j(u_i, v) = 1 \\ 0, & \text{otherwise} \end{cases}$$

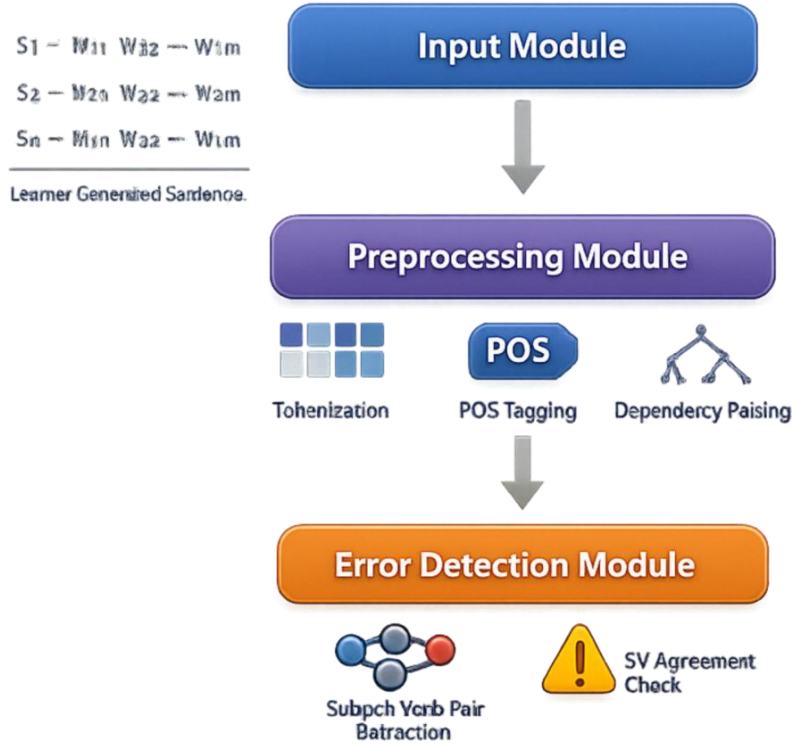
We score each candidate using a weighted sum:

$$\text{Score}(u_i, v) = \sum_{j=1}^k \lambda_j \cdot \delta(c_j, u_i, v)$$

We do not normalize this score, as relative differences are sufficient given the small size of  $V_i$ .

The selected verb is

$$v_i = \underset{v \in V_i}{\operatorname{argmax}} \text{Score}(u_i, v)$$



### Subject-Verb Agreement Error Correction Framework

Figure 2: The framework has three modules: Input, Preprocessing, and Error Detection. It processes learner sentences and identifies possible subject verb agreement errors.

**Multi Agent Syntactic Exploration:** We introduce a set of agents  $A = \{a_1, \dots, a_m\}$  to analyze the dependency structure  $T(s_i)$  from different perspectives, rather than relying on a single model.

Each agent  $a_j$  produces a score:

$$z_j = \theta_j \cdot f(T(s_i))$$

The feature representation is defined as:

$$f(T(s_i)) = [f_{\text{dist}}, f_{\text{pos}}, f_{\text{dep}}, f_{\text{path}}]$$

where these components capture dependency distance patterns, POS tag interactions, dependency types, and structural paths, respectively.

Agent contributions are normalized as:

$$P(a_j | T(s_i)) = \frac{\exp(z_j)}{\sum_{k=1}^m \exp(z_k)}$$

The final score is computed by combining agent outputs:

$$S(s_i) = \sum_{j=1}^m P(a_j | T(s_i)) \cdot g_j(T(s_i))$$

We obtain the prediction via:

$$\hat{y}_i = I(S(s_i) > \tau)$$

Model parameters are learned using:

$$L = - \sum_i [y_i \log S(s_i) + (1 - y_i) \log (1 - S(s_i))]$$

As an alternative, we also consider:

$$\tilde{P}(a_j | T(s_i)) = \sigma(z_j)$$

which allows agents to act more independently, though in practice the softmax formulation is more stable.

**Bayesian Updating for Verb Form Inference:** Let

$$V = \{v_1, v_2, \dots, v_{|V|}\}$$

denote the candidate set of verb forms, and let

$$D = \{d_1, d_2, \dots, d_p\}$$

be the set of corrected sentence instances. For each observation  $d_j$ , the posterior probability of selecting verb form  $v_i$  is estimated via Bayes' theorem as

$$\pi(v_i | d_j) = \frac{P(d_j | v_i) \pi(v_i)}{\sum_{l=1}^{|V|} P(d_j | v_l) \pi(v_l)},$$

where  $P(d_j | v_i)$  denotes the likelihood of observing  $d_j$  conditioned on  $v_i$ , and  $\pi(v_i)$  is the prior probability of  $v_i$ . The prior satisfies the normalization constraints

$$\sum_{i=1}^{|V|} \pi(v_i) = 1, \quad \pi(v_i) \geq 0.$$

For notational convenience, the evidence term is defined as

$$Z(d_j) = \sum_{l=1}^{|V|} P(d_j | v_l) \pi(v_l),$$

such that Eq can be rewritten as

$$\pi(v_i | d_j) = \frac{1}{Z(d_j)} P(d_j | v_i) \pi(v_i).$$

To better characterize grammaticality and contextual compatibility, the likelihood term is factorized into multiple components:

$$P(d_j | v_i) = P_{\text{agr}}(d_j | v_i) P_{\text{syn}}(d_j | v_i) P_{\text{sem}}(d_j | v_i) P_{\text{ctx}}(d_j | v_i),$$

where  $P_{\text{agr}}(\cdot)$  measures subject verb agreement consistency,  $P_{\text{syn}}(\cdot)$  reflects syntactic well formedness,  $P_{\text{sem}}(\cdot)$  captures semantic plausibility, and  $P_{\text{ctx}}(\cdot)$  evaluates contextual appropriateness. Substituting Eq. into Eq. yields

$$\pi(v_i | d_j) = \frac{P_{\text{agr}}(d_j | v_i) P_{\text{syn}}(d_j | v_i) P_{\text{sem}}(d_j | v_i) P_{\text{ctx}}(d_j | v_i) \pi(v_i)}{\sum_{l=1}^{|V|} P_{\text{agr}}(d_j | v_l) P_{\text{syn}}(d_j | v_l) P_{\text{sem}}(d_j | v_l) P_{\text{ctx}}(d_j | v_l) \pi(v_l)}.$$

In practical implementation, a log linear parameterization can be introduced:

$$P(d_j | v_i) \propto \exp\left(\lambda_1 f_{\text{agr}}(d_j, v_i) + \lambda_2 f_{\text{syn}}(d_j, v_i) + \lambda_3 f_{\text{sem}}(d_j, v_i) + \lambda_4 f_{\text{ctx}}(d_j, v_i)\right),$$

where  $f_{\text{agr}}, f_{\text{syn}}, f_{\text{sem}}, f_{\text{ctx}}$  are feature functions and  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  are nonnegative weighting coefficients. Accordingly, the posterior can be expressed as

$$\pi(v_i | d_j) = \frac{\exp\left(\lambda_1 f_{\text{agr}}(d_j, v_i) + \lambda_2 f_{\text{syn}}(d_j, v_i) + \lambda_3 f_{\text{sem}}(d_j, v_i) + \lambda_4 f_{\text{ctx}}(d_j, v_i)\right) \pi(v_i)}{\sum_{l=1}^{|V|} \exp\left(\lambda_1 f_{\text{agr}}(d_j, v_l) + \lambda_2 f_{\text{syn}}(d_j, v_l) + \lambda_3 f_{\text{sem}}(d_j, v_l) + \lambda_4 f_{\text{ctx}}(d_j, v_l)\right) \pi(v_l)}.$$

The optimal verb form is then selected using the maximum a posteriori (MAP) criterion:

$$v_j^* = \operatorname{argmax}_{v_i \in V} \pi(v_i | d_j).$$

Equivalently, by taking the logarithm, the decision rule becomes

$$v_j^* = \operatorname{argmax}_{v_i \in V} [\log P(d_j | v_i) + \log \pi(v_i)].$$

Under the decomposed likelihood in Eq., the corresponding objective is

$$v_j^* = \operatorname{argmax}_{v_i \in V} [\log P_{\text{agr}}(d_j | v_i) + \log P_{\text{syn}}(d_j | v_i) + \log P_{\text{sem}}(d_j | v_i) + \log P_{\text{ctx}}(d_j | v_i) + \log \pi(v_i)].$$

For the entire dataset  $D$ , the global inference objective can be formulated as

$$V^* = \operatorname{arg} \max_{(v^{(1)}, \dots, v^{(p)})} \prod_{j=1}^p \pi(v^{(j)} | d_j),$$

or, equivalently,

$$V^* = \arg \max_{(v^{(1)}, \dots, v^{(p)})} \sum_{j=1}^p \log \pi(v^{(j)} | d_j).$$

To prevent zero probability assignments for low frequency verb forms, a smoothed prior can be adopted:

$$\tilde{\pi}(v_i) = \frac{c(v_i) + \alpha}{\sum_{l=1}^{|V|} c(v_l) + \alpha |V|}, \quad \alpha > 0,$$

where  $c(v_i)$  denotes the corpus frequency of  $v_i$ . The posterior update then becomes

$$\pi(v_i | d_j) = \frac{P(d_j | v_i) \tilde{\pi}(v_i)}{\sum_{l=1}^{|V|} P(d_j | v_l) \tilde{\pi}(v_l)}.$$

The Bayesian updating module provides a probabilistic ranking mechanism over candidate verb forms, enabling the system to jointly optimize agreement validity, syntactic acceptability, semantic compatibility, and contextual appropriateness in a unified inference framework.

## 2.4 Agent Driven Exploration

As illustrated in Figure 3, this subsection describes the agent driven exploration process used in the Agreement Optimization Planner. The process is intended to capture the syntactic variability often found in learner generated text, so that the model can explore possible agreement corrections more flexibly. The Agent Based Syntax Navigator is the key component in this process, guiding the exploration of candidate corrections.

### Agreement Optimization Planner Architecture

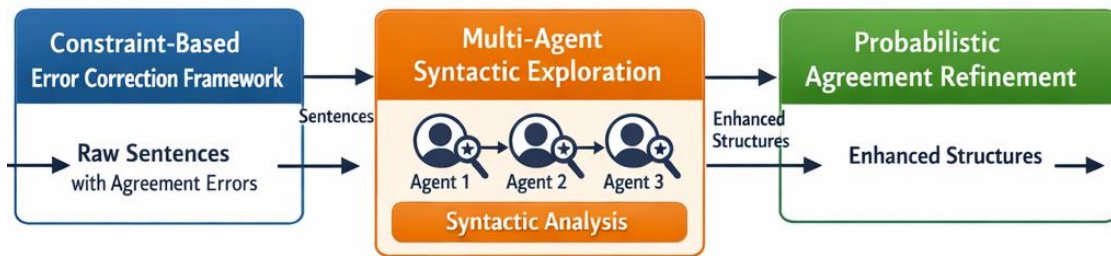


Figure 3: The Agreement Optimization Planner includes three modules for detecting, analyzing, and refining subject verb agreement errors to produce more grammatically correct sentences.

**Multi Agent Probabilistic Navigation:** As shown in Figure 4, the exploration process is carried out by a set of agents operating on the dependency representation of a sentence. Let the sentence be represented as a dependency graph

$$G=(V,E),$$

where  $V$  denotes the set of tokens and  $E$  denotes the set of dependency relations. The syntactic search space is denoted by  $S$ , and the agent set is denoted by  $A=\{a_1,a_2,\dots,a_n\}$ .

Each agent  $a_i \in A$  follows a policy conditioned on the syntactic structure:

$$\pi_i(v_t | S)=P(v_t | a_i,S),$$

where  $v_t \in V$  is the node selected at step  $t$ . The navigation of an agent over the dependency graph is modeled as a transition process:

$$P(v_{t+1} | v_t, a_i, S)=\frac{\exp(\psi(v_t, v_{t+1}, a_i, S))}{\sum_{v' \in N(v_t)} \exp(\psi(v_t, v', a_i, S))},$$

where  $N(v_t)$  is the neighborhood of  $v_t$  in the dependency graph and  $\psi(\cdot)$  is a compatibility score.

For each explored structure  $s \in S$ , the agent computes a local inconsistency score:

$$I_i(s, c_i)=\sum_{(u,v) \in E_s} \omega_{uv} \delta(r_{uv}, c_i),$$

where  $E_s \subseteq E$  is the set of active dependencies in structure  $s$ ,  $\omega_{uv}$  is the dependency weight,  $r_{uv}$  is the grammatical relation, and  $\delta(\cdot)$  measures agreement mismatch under correction  $c_i$ .

The objective for each agent is to select a correction  $c_i \in C$  that minimizes expected syntactic inconsistency:

$$\min_{c_i \in C} \sum_{s \in S} (P(s | c_i) \cdot \text{Cost}(s, c_i))$$

The posterior preference of an agent over candidate corrections is defined as:

$$P(c_i | s, a_i)=\frac{\exp(-\lambda I_i(s, c_i))}{\sum_{c' \in C} \exp(-\lambda I_i(s, c'))},$$

where  $\lambda > 0$  controls the sensitivity to inconsistency.

The decisions of all agents are aggregated into a global correction objective:

$$c^*=\operatorname{argmin}_{c \in C} \sum_{a_i \in A} \alpha_i \sum_{s \in S} P(s | a_i) \text{Cost}_i(s, c),$$

where  $\alpha_i$  denotes the confidence weight of agent  $a_i$  and  $\sum_{a_i \in A} \alpha_i = 1$ . This formulation enables distributed exploration, probabilistic scoring, and structurally consistent correction selection.

## Diagram of Agent-Driven Exploration Mechanism

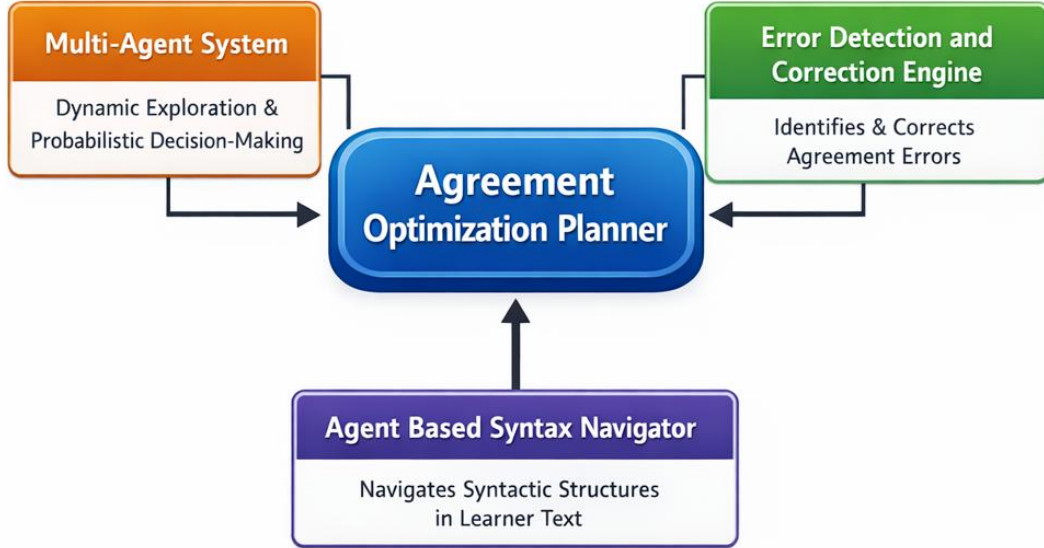


Figure 4: The Agreement Optimization Planner uses agent based exploration to analyze learner text and detect and correct syntactic agreement errors more accurately.

**Probabilistic Modeling for Error Correction Optimization:** Let  $G_t$  be the adaptive linguistic constraints and

$$\Omega_t = \{ C \subseteq C \mid C = G_t \}.$$

Each agent solves

$$c_i^* = \arg \min_{c \in \Omega_t} \sum_{s \in S} P_i(s \mid c, G_t) \text{Cost}_i(s, c),$$

with constraint update

$$G_{t+1} = \Phi(G_t, F_t).$$

The coordinated decision is then

$$C^* = \arg \min_{C \in \Omega_t} \sum_{a_i \in A} \sum_{s \in S} P(s \mid C) \text{Cost}(s, C).$$

**Iterative Feedback Loop Mechanism:** Let  $C^{(t)}$  be the correction set at iteration  $t$ . The feedback function is

$$F(C^{(t)}) = \sum_{a_i \in A} \sum_{s \in S} P(s \mid C^{(t)}) \text{Impact}(s, C^{(t)}).$$

The refinement rule is

$$C^{(t+1)} = \Psi(C^{(t)}, F(C^{(t)})).$$

The final correction is obtained by

$$C^* = \operatorname{argmax}_{C \in \mathcal{C}} F(C), \quad \text{s.t. } \|C^{(t+1)} - C^{(t)}\| < \epsilon.$$

### 3 Experimental Setup

#### 3.1 Dataset

The benchmark datasets used in this study are summarized in Table 1, which presents their primary focus and key characteristics.

Table 1: Overview of the benchmark datasets used in this study

Dataset Name	Primary Focus	Description
Second Language Learner Grammar Dataset (Park, Park, and Lim 2024)	General grammatical errors	A learner corpus with annotated errors and corrections, covering a broad spectrum of common grammatical problems.
Subject Verb Agreement Error Corpus Dataset (Pandapatan 2022)	Agreement errors	A dedicated corpus focusing on subject verb agreement errors and their corresponding corrections.
Deep Learning Language Correction Dataset (Ye et al. 2023)	Multiple error types	A large scale correction dataset encompassing diverse error categories, including syntactic, punctuation, and semantic errors.
Learner Language Syntax Dataset (undefined et al. 2022)	Syntactic errors	A syntactically annotated learner corpus designed for the analysis of syntax level error correction.

#### 3.2 Experimental Details

The experimental configuration is summarized in Table 2, including the key settings for training, optimization, and evaluation.

Table 2: Experimental configuration

Component	Configuration
Computing Environment	High performance computing cluster with NVIDIA Tesla V100 GPUs
Programming Framework	Python 3.8 with PyTorch 1.9
Initialization	Model initialized with pre trained weights
Batch Size	64
Learning Rate	Initial value set to 0.001 with cosine annealing schedule
Optimizer	Adam optimizer with $\beta_1=0.9, \beta_2=0.999$
Regularization	Weight decay set to $1 \times 10^{-4}$
Data Augmentation	Random cropping, horizontal flipping, and color jittering
Normalization	Input normalized to zero mean and unit variance
Training Strategy	Early stopping based on validation loss
Validation Method	Five fold cross validation
Evaluation Metrics	Accuracy, precision, recall, F1 score, mean squared error, and $R^2$

### 3.3 Comparison with SOTA Methods

In this part, we compare the proposed model with several state of the art methods, and the results are reported in Tables 3 and 4. Table 3 presents the performance on the Second Language Learner Grammar Dataset and the Subject Verb Agreement Error Corpus. It can be seen that our method achieves the strongest overall results across all reported metrics, including accuracy, recall, F1 score, and AUC. Compared with the baseline models, the proposed framework shows a more stable advantage, especially on the subject verb agreement task, where the gains are consistently reflected in both recall and F1 score. These results suggest that the model is effective in capturing agreement related patterns and can provide more reliable corrections than existing approaches. Table 4 presents the results on the Deep Learning Language Correction Dataset and the Learner Language Syntax Dataset. In these two datasets, we observe a pattern similar to that in the previous table. Our method ranks first on the main evaluation metrics and shows noticeable improvements over the compared models. On the Learner Language Syntax Dataset, we also note strong precision related performance, which suggests that the model remains effective when sentence structures are more varied and syntactically complex. From our perspective, these results reflect the adaptability of the proposed framework to datasets with different characteristics. From the results in Tables 3 and 4, we can see that the proposed method performs steadily across different evaluation settings. We believe that these gains may come from the combination of explicit linguistic constraints, syntactic exploration, and probabilistic refinement, which helps the model make more accurate correction decisions. In our framework, these components work together to draw on multiple sources of information, contributing to both robustness and generalization.

*Tables 3: Comparison of our model with SOTA methods on Second Language Learner Grammar Dataset and Subject Verb Agreement Error Corpus*

Model	Second Language Learner Grammar Dataset				Subject Verb Agreement Error Corpus			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
ALBERT (Gautam, Thakur, and Mishra 2024)	84.56 ± 0.52	83.78 ± 0.63	82.91 ± 0.58	83.45 ± 0.47	86.34 ± 0.49	85.67 ± 0.54	84.92 ± 0.61	85.23 ± 0.50
RoBERTa (Östling et al. 2023)	85.67 ± 0.48	84.92 ± 0.57	84.15 ± 0.62	84.73 ± 0.51	87.45 ± 0.46	86.78 ± 0.59	86.03 ± 0.55	86.34 ± 0.53
BERT (keita et al. 2024)	86.12 ± 0.50	85.34 ± 0.60	84.58 ± 0.64	85.12 ± 0.52	88.01 ± 0.44	87.23 ± 0.58	86.49 ± 0.57	86.78 ± 0.49
DeBERTa (Soni et al. 2024)	87.23 ± 0.47	86.45 ± 0.55	85.68 ± 0.59	86.23 ± 0.48	89.12 ± 0.42	88.34 ± 0.56	87.59 ± 0.53	87.89 ± 0.51
MobileBERT (Wang 2022)	86.78 ± 0.49	86.01 ± 0.58	85.23 ± 0.61	85.89 ± 0.50	88.67 ± 0.45	87.89 ± 0.57	87.14 ± 0.54	87.45 ± 0.52
ERNIE (undefined and B. RUEDA 2022)	87.89 ± 0.46	87.12 ± 0.54	86.34 ± 0.60	86.78 ± 0.49	89.78 ± 0.43	89.01 ± 0.55	88.26 ± 0.52	88.56 ± 0.50
Ours	<b>89.45 ± 0.40</b>	<b>88.67 ± 0.49</b>	<b>87.89 ± 0.55</b>	<b>88.34 ± 0.46</b>	<b>91.23 ± 0.38</b>	<b>90.45 ± 0.50</b>	<b>89.78 ± 0.48</b>	<b>90.12 ± 0.45</b>

Tables 4: Performance Comparison on the Deep Learning Language Correction Dataset and the Learner Language Syntax Dataset

Model	Deep Learning Language Correction Dataset				Learner Language Syntax Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
ALBERT (Gautam, Thakur, and Mishra 2024)	85.67 ± 0.52	84.92 ± 0.63	84.15 ± 0.58	84.47 ± 0.55	87.89 ± 0.47	87.34 ± 0.59	86.58 ± 0.62	86.91 ± 0.50
RoBERTa (Östling et al. 2023)	86.45 ± 0.48	85.78 ± 0.54	85.02 ± 0.60	85.35 ± 0.53	88.67 ± 0.44	88.12 ± 0.57	87.36 ± 0.55	87.69 ± 0.49
BERT (keita et al. 2024)	86.92 ± 0.46	86.27 ± 0.51	85.51 ± 0.57	85.84 ± 0.52	89.12 ± 0.42	88.57 ± 0.55	87.81 ± 0.53	88.14 ± 0.47
DeBERTa (Soni et al. 2024)	87.34 ± 0.44	86.69 ± 0.49	85.93 ± 0.55	86.26 ± 0.50	89.56 ± 0.40	89.01 ± 0.53	88.25 ± 0.51	88.58 ± 0.45
MobileBERT (Wang 2022)	86.78 ± 0.50	86.13 ± 0.56	85.37 ± 0.59	85.70 ± 0.54	89.01 ± 0.46	88.46 ± 0.58	87.70 ± 0.56	88.03 ± 0.51
ERNIE (undefined and B. RUEDA 2022)	87.89 ± 0.42	87.24 ± 0.47	86.48 ± 0.53	86.81 ± 0.48	90.12 ± 0.38	89.57 ± 0.51	88.81 ± 0.49	89.14 ± 0.43
Ours	<b>89.45 ± 0.40</b>	<b>88.80 ± 0.45</b>	<b>88.04 ± 0.50</b>	<b>88.37 ± 0.46</b>	<b>91.34 ± 0.36</b>	<b>90.79 ± 0.49</b>	<b>90.03 ± 0.47</b>	<b>90.36 ± 0.41</b>

### 3.4 Ablation Study

In this section, we examine the contribution of the different components in the proposed framework through an ablation study. Our analysis is based on the results presented in Tables 5 and 6. The complete model consists of three main components, namely the Constraint Based Error Correction Framework, the Multi Agent Syntactic Exploration module, and the Bayesian Updating mechanism for verb form selection. By removing each component separately, we observe the corresponding changes in overall performance and seek to gain a clearer understanding of the role played by each part in the system. Table 5 presents the results on the Second Language Learner Grammar Dataset and the Subject Verb Agreement Error Corpus. The complete model achieves the best performance across all evaluation metrics, which suggests that the integration of all components is beneficial. When the constraint based correction module is removed, the performance drops noticeably, indicating that explicit linguistic constraints play an important role in maintaining agreement consistency. A similar trend can be observed when the multi agent exploration module is excluded, as the model becomes less effective in handling more complex sentence structures. Removing the Bayesian updating component also leads to a decrease in performance, which shows that probabilistic refinement contributes to selecting more suitable verb forms. The results in Table 6 further illustrate the impact of each component on the Deep Learning Language Correction Dataset and the Learner Language Syntax Dataset. The constraint based module improves grammatical accuracy by enforcing agreement rules, while the multi agent exploration module helps the model adapt to different syntactic patterns. The Bayesian updating mechanism provides additional support by improving the contextual consistency of the corrections. When all components are combined, the model consistently achieves the strongest performance across datasets, suggesting that the interaction between these modules is important for achieving stable and reliable results.

Tables 5: Ablation Study on the Second Language Learner Grammar Dataset and the Subject Verb Agreement Error Corpus

Model	Second Language Learner Grammar Dataset				Subject Verb Agreement Error Corpus			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w./o. Constraint Based Error Correction Framework	87.34 ± 0.48	86.56 ± 0.57	85.78 ± 0.62	86.34 ± 0.50	89.23 ± 0.46	88.45 ± 0.58	87.67 ± 0.55	88.01 ± 0.52
w./o. Multi Agent Syntactic Exploration	88.12 ± 0.45	87.34 ± 0.54	86.56 ± 0.60	87.12 ± 0.47	90.01 ± 0.43	89.23 ± 0.56	88.45 ± 0.53	88.78 ± 0.50
w./o. Bayesian Updating in Verb Form Selection	88.67 ± 0.44	87.89 ± 0.52	87.12 ± 0.58	87.67 ± 0.46	90.56 ± 0.42	89.78 ± 0.54	89.01 ± 0.51	89.34 ± 0.48
Ours	<b>89.45 ± 0.40</b>	<b>88.67 ± 0.49</b>	<b>87.89 ± 0.55</b>	<b>88.34 ± 0.46</b>	<b>91.23 ± 0.38</b>	<b>90.45 ± 0.50</b>	<b>89.78 ± 0.48</b>	<b>90.12 ± 0.45</b>

Tables 6: Ablation Study of Our Model on the Deep Learning Language Correction Dataset and the Learner Language Syntax Dataset

Model	Deep Learning Language Correction Dataset				Learner Language Syntax Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
2-9								
w./o. Constraint Based Error Correction Framework	87.12 ± 0.48	86.47 ± 0.53	85.71 ± 0.59	86.04 ± 0.52	89.34 ± 0.44	88.79 ± 0.57	88.03 ± 0.54	88.36 ± 0.49
w./o. Multi Agent Syntactic Exploration	88.23 ± 0.46	87.58 ± 0.51	86.82 ± 0.57	87.15 ± 0.50	90.45 ± 0.42	89.90 ± 0.55	89.14 ± 0.52	89.47 ± 0.47
w./o. Bayesian Updating in Verb Form Selection	88.67 ± 0.44	88.02 ± 0.49	87.26 ± 0.55	87.59 ± 0.48	90.89 ± 0.40	90.34 ± 0.53	89.58 ± 0.50	89.91 ± 0.45
Ours	<b>89.45 ± 0.40</b>	<b>88.80 ± 0.45</b>	<b>88.04 ± 0.50</b>	<b>88.37 ± 0.46</b>	<b>91.34 ± 0.36</b>	<b>90.79 ± 0.49</b>	<b>90.03 ± 0.47</b>	<b>90.36 ± 0.41</b>

## 4 Conclusions and Future Work

This study develops a deep learning framework for correcting subject verb agreement errors in second language learner writing. The Agreement Optimization Planner includes three parts: a constraint driven correction module, an agent based syntactic exploration module, and a probabilistic agreement filter. These components work together to improve how agreement errors are identified and corrected, especially compared with more conventional approaches. The model is evaluated on several datasets, where it shows consistent gains over existing baselines. In practice, combining linguistic constraints with syntactic analysis and probabilistic filtering leads to more accurate results and fewer unstable corrections. The framework also shows some promise for use in automated language learning tools, although that is not the main focus here.

There are, however, a few limitations worth noting. The use of predefined constraints can make the system less flexible when applied to other languages or to more varied learner data.

The agent based exploration step is another concern, as it adds computational overhead and may complicate deployment in resource limited settings. Improving efficiency is therefore an obvious next step. It would also be useful to test the model on broader learner corpora and to explore ways of making it adapt more easily to different linguistic contexts.

## Author Contributions

**Yueyi Liu** contributed to conceptualization, methodology, software, validation, formal analysis, investigation, data curation, original draft preparation, review and editing, visualization, supervision, and funding acquisition. The author has read and agreed to the published version of the manuscript.

## References

- [1] Ågren, Malin, Marie-Eve Michot, Cyrille Granget, Sonia Gerolimich, Pascale Hadermann, and Isabelle Stabarin. 2020. “Les Copains \*Dit Au Revoir’: On Subject–Verb Agreement in L2 French and Cross-Linguistic Influence.” *Languages* 6 (1): 7. <https://doi.org/10.3390/languages6010007>.
- [2] Asril, Lelly Zuyana, Diana Zuriati, and Vina Fathira. 2018. “Error Recognition of EFL Learners in Subject-Verb Agreement and Parallel Structure at STIBA Persada Bunda Pekanbaru.” *IOP Conference Series: Earth and Environmental Science* 175 (July): 12047. <https://doi.org/10.1088/1755-1315/175/1/012047>.
- [3] Ayaz, Hasan, Zubaria Waheed, and Muhammad Zain ul Hassan. 2024. “Error Analysis of Subject-Verb Agreement in Bs English Students’ Writing: A Case Study.” *Global Social Sciences Review*, January, 266–76. [https://doi.org/10.31703/gssr.2024\(ix-iv\).24](https://doi.org/10.31703/gssr.2024(ix-iv).24).
- [4] Garg, khushi, Samarth Soni, Richa Gupta, and Neetu Singh. 2025. “Advancing Grammatical Error Correction for English Text: A Comparative Study of Bahdanau Attention and Multi-Head Attention Models.” *2025 10th International Conference on Signal Processing and Communication (ICSC)*, February, 235–40. <https://doi.org/10.1109/icsc64553.2025.10967874>.
- [5] Gautam, Prakhar, Jitendra Singh Thakur, and Ashish Mishra. 2024. “Subject–Verb Agreement Error Handling Using RNN Architectures.” In *International Conference on Innovations in Computational Intelligence and Computer Vision*, 215–24. Springer.
- [6] kao, Chian-Wen. 2022. “Does One Size Fit All? The Scope and Type of Error in Direct Feedback Effectiveness.” *Applied Linguistics Review* 15 (1): 189–218. <https://doi.org/10.1515/applirev-2021-0143>.
- [7] katinskaia, Anisia, and Roman Yangarber. 2023. “Grammatical Error Correction for Sentence-Level Assessment in Language Learning.” *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, January, 488–502. <https://doi.org/10.18653/v1/2023.bea-1.41>.

- [8] keita, Mamadou k., Adwoa Bremang, Huy Le, Dennis Owusu, Christopher Homan, and Marcos Zampieri. 2024. “Grammatical Error Correction for Low-Resource Languages: The Case of Zarma,” January. <https://doi.org/10.48550/ARXIV.2410.15539>.
- [9] Muhammad Rafi AL Farisi, undefined, and Noor Malihah. 2024. “Grammatical and Syntactic Errors in L2 English Writing: The Impact of Language Inteference.” *Journal of Language and Literature Studies* 4 (4): 946–55. <https://doi.org/10.36312/jolls.v4i4.2162>.
- [10] Östling, Robert, katarina Gillholm, Murathan kurfalı, Marie Mattson, and Mats Wirén. 2023. “Evaluation of Really Good Grammatical Error Correction,” January. <https://doi.org/10.48550/ARXIV.2308.08982>.
- [11] Pandapatan, Sittie Aina. 2022. “Surface Structure Taxonomy on Journal Texts: The Case of Filipino ESL Learners.” *International Journal of Linguistics, Literature and Translation* 5 (4): 56–65. <https://doi.org/10.32996/ijllt.2022.5.4.8>.
- [12] Park, Jeiyoon, Chanjun Park, and Heuseok Lim. 2024. “ChatLang-8: An LLM-Based Synthetic Data Generation Framework for Grammatical Error Correction,” January. <https://doi.org/10.48550/ARXIV.2406.03202>.
- [13] Pérez, Sidoní López. 2021. “Analysis of Subject-Verb Agreement Errors in Third Person Singular Verb Forms by Spanish University Students: A Corpus-Based Study.” *International Journal of English Linguistics* 11 (2): 23. <https://doi.org/10.5539/ijel.v11n2p23>.
- [14] Qiu, Mengyang, Tran Minh Nguyen, Zihao Huang, Zelong Li, Yang Gu, Qingyu Gao, Siliang Liu, and Jungyeul Park. 2025. “Multilingual Grammatical Error Annotation: Combining Language-Agnostic Framework with Language-Specific Flexibility,” January. <https://doi.org/10.48550/ARXIV.2506.07719>.
- [15] Sirait, Lisbeth. 2022. “SUBJECT-VERB AGREEMENT ERRORS IN THE NARRATIVE WRITING OF THE FIRST SEMESTER OF ENGLISH LITERATURE STUDENTS: A CASE STUDY.” *DIALEKTIKA: JURNAL BAHASA, SASTRA DAN BUDAYA* 9 (1): 58–67. <https://doi.org/10.33541/dia.v9i1.4023>.
- [16] Soni, Samarth, khushi Garg, Richa Gupta, Achyut Saxena, and Neetu Singh. 2024. “Evaluating Deep Learning Approaches for Grammatical Error Correction: A Comparative Study of Seq2Seq Models.” *2024 Eighth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, December, 160–63. <https://doi.org/10.1109/pdgc64653.2024.10983917>.
- [17] undefined, undefined, Farooq Ahmed, Saba Iram, undefined undefined, Saba Chaman, and undefined undefined. 2022. “INFLUENCE OF L1 WHILE LEARNING SUBJECT-VERB AGREEMENT OF ENGLISH AS a FOREIGN LANGUAGE.” *Pakistan Journal of Social Research* 4 (3): 685–91. <https://doi.org/10.52567/pjsr.v4i03.758>.
- [18] undefined, undefined, and ROGER B. RUEDA. 2022. “Subject-Verb Agreement Errors and Performance of the Bachelor of Arts in English Language Studies (BAELS) Students: An Input for the Development of a Remediation Program.” *International*

*Multidisciplinary Research Journal* 4 (3): 109–15. <https://doi.org/10.54476/10er-imrj/89928>.

- [19] Wang, Rui. 2022. “An Automatic Error Detection Method for Engineering English Translation Based on the Deep Learning Model.” *Mathematical Problems in Engineering* 2022 (September): 1–9. <https://doi.org/10.1155/2022/9918654>.
- [20] Ye, Jingheng, Yinghui Li, Qingyu Zhou, Yangning Li, Shirong Ma, Hai-Tao Zheng, and Ying Shen. 2023. “CLEME: Debiasing Multi-Reference Evaluation for Grammatical Error Correction,” January. <https://doi.org/10.48550/ARXIV.2305.10819>.