Automated Grammatical Error Correction: a Comprehensive Review

Sadaf Abdul Rauf, Ramsha Saeed, Noorulain Saleem Khan, Kiran Habib, Pernia Gabrail and Fareeha Aftab

Fatima Jinnah Women University, The Mall, Rawalpindi, Pakistan

Email: {sadaf.abdulrauf, noorulain93, kirnrana21.10, pernia.gabrial, fareeha.ch5}@gmail.com, ramsha.saeed93@yahoo.com

Abstract

Automatic Grammatical Error Correction is one of the most challenging and continuously evolving areas of linguistics which aims at automatically detecting and correcting the grammatical errors in the text. Such systems are specifically helpful for non-native speakers or learners of a certain language. Grammatical Error Correction has been the focus of latest research and various techniques have been employed to cater the grammatical error categories of different languages. English being one of the most widely spoken languages across the world has been an area of research for decades. Extensive work has been done for developing better systems which could analyse the syntactic structure of English text and detect and correct grammatical errors made by the writers. The objective of this article is to present a review of state of the art research done for automated grammatical error correction of English language. The article discusses some of the most commonly encountered error categories in English grammar with examples. The most widely used approaches in the literature have been discussed and the respected works summarized. Conclusively, all the works are analysed and compared on the basis of the strengths of the used techniques, the number of error categories considered in each work and the metrics used for their evaluation.

Keywords: Grammatical Error Correction, Computational Linguistics, Rule Based, Machine Learning

1. Introduction

In the past several years, the area of computational linguistics has seen a remarked improvement in the field of Grammatical Error Correction (GEC). This deals with automatic identification and correction of grammatical errors present in some written text [1]. Grammatical error identification deals with identifying the grammatical errors in the given sentence while grammatical error correction deals with identifying the errors and also suggesting and applying the corresponding correction. Grammar checking is a very complex task as natural languages do not have any specific syntax like those of computer programming languages. Even though complete rules for a formal grammar of natural language may be written, formal grammar checker may fail to deal with a number of grammatical exceptions of a language in real usage. The most important component of grammar error correctors is a lexicon containing all the words of a language as well as part of speech of each word. For each sentence detected in the text the program finds each word in the dictionary, parses the sentence according to its grammatical structure, and then detects errors.

Mostly, grammatical error checkers are a part of some larger applications; one of the best examples being Microsoft Word Processor. Grammar checker has been a part of Word since its release in 1997 [2]. However, stand-alone applications of grammar checkers also exist such as 'Language Tool' 1. Such types of systems are useful in terms of providing learning aid to different foreign language learners across the world [3-8].

The research in the field of GEC dates back to 1970s when the first grammar checker called 'Writer's Workbench' was developed and was included in the UNIX systems as a set of writing tools [9]. Recently, this field has intrigued many researchers with a primary focus on correction of grammatical errors in the English language to help the learners of English as a second language (ESL) [1, 4, 3, 8, 10, 11, and 12]. For this purpose, four shared tasks took place in the past few years namely: 'Helping Our Own' (HOO) held in 2011 [13] and 2012 [14], and the CoNLL shared tasks organized in 2013 and 2014 [15, 16]. All of these provided an annotated corpus of learner text and a test set hence promoting more research in the area. The most reputable CoNLL shared task-2014,which came up as an extension of CoNLL shared task-2013, covers a number of errors for ESL including mistakes in prepositions, verbs, nouns and in the use of articles.

Normally, three approaches are used for detecting and correcting the grammatical errors: knowledge engineering/rule based approach, machine learning approach and hybrid approach. Rule based approach is based on writing all the grammatical rules, machine learning emphasizes on making the system learn from the training data and predict the result while hybrid approach uses a combination of the aforementioned approaches. All of these approaches have their own advantages and disadvantages. A set of rules may not be able to cater grammatical exceptions of a language. On the other hand, the training data used for machine learning approach may consist of text annotated with grammatical errors and could be noisy. However, hybridization of these approaches may overcome the challenges of both, leading to improved results.

The paper is organized as follows. Section 2 presents frequently encountered error categories; Section 3 gives the survey of the major existing approaches used for GEC. Section 4 briefs the work done on GEC with reference to speech, Section 5 discusses the metrics used for evaluation in all the reviewed works, Section 7 outlines the error categories targeted in all the works and finally the article concludes with the Conclusions section.

2. Grammatical Errors in English Language

Table 1 presents some of the categories of errors commonly made by English language learners. Row 2 of the table presents common Subject Verb agreement errors which occur when there is disagreement between subject

DOI: http://dx.doi.org/10.24949njes.v10i2.219

and verb in number or person, for example, in the erroneous sentence "He eat an apple daily", the verb doesn't agree with the subject, the correct formation should be "He eats an apple daily". Table 2 shows the distribution of errors in KJ corpus as reported by [17] and Table 3 shows the distribution of errors in CoNLL-2013 corpus as reported by [18]. According to the two tables article errors are the most common errors followed by noun and prepositional errors.

3. Grammatical Error Correction Approaches

The approaches typically used for GEC are rule based, machine learning and hybrid. Different methods and sub-techniques used by the researchers have been tabulated in Table 4.

In this section we present detailed overview of some of the major techniques that have been used so far for detecting and correcting grammatical errors.

3.1. Rule Based Approaches

Rule based systems rely on the knowledge of manually written grammatical rules of a certain language such as syntax, semantics, morphology etc. Writing all the rules for languages is time consuming and laborious. Nagata et al., [7] put forward a procedure for error detection to distinguish between mass and count nouns. The corpus used was British national corpus. In the first step the system learned a list of decisions from a training data. Then, the corpus to be used was augmented by feedback i.e. it included writings of Japanese learners that were checked by an English teacher and hence was called a feedback corpus. Lastly, it detected the errors by applying rules that it had previously learned. The experiment results proved that the system performed better than any other using the same technique at that time with 72%precision.

Sidorov et al., [19] presented a rule based GEC system for the English language which took part in CoNLL 2013 shared tasks and used Nus Corpus of Learner English (NUCLE). The system uses simple correction rules without using any complex linguistic resources for five error types noun number, subject-verb agreement, verb form, article/determiner and choice of preposition and uses syntactic n-grams in some cases. Due to simplicity and fewer resources, the system could not obtain high scores having 8.13% precision, 12.42% recall and F1 measure of 9.83%, and can be regarded as the baseline system; however the research describes the situation where rule based systems work well.

3.2. Machine Learning Approaches

The alternative to rule based approach is the machine learning approach in which annotated corpora are created to train the system instead of hand-crafted rules. It is a type of artificial intelligence which enables the computer to iteratively learn patterns from data (also called training data) and make predictions on it. The training process leads to the creation of a model which, when exposed to some new data called the test data, has to make predictions on it. The machine learning algorithms consider a number of textual and contextual features of a language at many levels.

Two of the most commonly used machine learning methods are supervised learning and unsupervised learning. In supervised learning, labelled examples are used to train the algorithm i.e. the examples where the desired outputs are known. For example, an input example sentence of a language could either be labelled as erroneous or correct. The learning algorithm is exposed to a set of examples with the corresponding actual outputs to make it learn features from those input examples. The training of the algorithm is stopped when the error of the predicted output over the training data is minimized by comparing the predicted output with the desired output and the model is modified accordingly. The model is then used to predict the output from unlabelled set of examples in the test data. Supervised learning is usually used in applications where probable future events need to be anticipated using historic data. Classification, regression and gradient boosting are some of the methods of supervised machine learning.

In unsupervised learning, the training dataset is not labelled with the desired outputs. The algorithm has to explore the data and find out the distinguishing features between different examples itself. For example, it can be used to separate textual content from each other on the basis of the topic, by identifying the attributes similar within a set of texts and different from the other text. Singular value decomposition, k-means clustering, self-organizing maps and nearest neighbour mapping are some of the common techniques of unsupervised learning.

Here, we present most widely used machine learning approaches employed by different researchers for GEC.

3.2.1. Classification Approach

Classification based approach is the dominant approach in which a classifier is trained to predict the most probable correction from the set of possible corrections using some features of the context of the target sentence. If the output of the prediction is different from the original word used by the writer, correction is applied and the word is replaced by the predicted word. Classification approach can cater each specific error category by training a separate classifier for each error type.

The authors in [10] used decision tree approach for detection and correction of mainly two types of errors i.e. determiners and choice of prepositions. They used 5-gram language model trained on English Gigaword corpus. The system worked with an average accuracy of 86.07%. Tetreault & Chodorow [12] used Maximum Entropy model to train the classifier for identifying prepositional errors in non-native English writings, they used the Meta metrics Lexile Corpus from Google. The proposed structure gave 84% precision and almost 19% level of precision when applied to a huge amount of data from student essays.

In 2010, Han, Tetreault, Lee, & Ha [11] carried out further research on prepositional errors in written texts by learners of the English language. The systems were trained on a huge corpus named Chungdahm English Learner Corpus using Maximum Entropy Model that featured grammatically annotated error sets that were written by learners of English as a foreign language (EFL). Preposition replacement errors, when tested, showed results that were 93.3% accurate and had recall rate of 14.8%. Secondly, their experiments showed that their model with error-annotated data performed better than those that are trained on edited text written by native English speakers.

Jia, Wang, & Zhao [21] demonstrated that GEC system could be converted into a multiclass classification problem and with the help of maximum entropy model it could be implemented as a single model system in which one model is used for all types of the errors. Maximum entropy model had been used as a classifier to attain the types of the errors and rules are applied to make corrections. The system was trained using NUCLE dataset. Multiple error types are considered such as determiners, prepositions, modal verbs, noun number, verb number, verb tense etc. F1 score (the harmonic mean of precision and recall) of 17.13% was achieved by the system.

An automated model for error correction with two main features was presented by Gamon, et al. [3]. One of the features was the training of classifiers through a machine learning approach. They were trained on large scale native data combined with a language model in order to improve the precision of the recommended error corrections. Secondly, the model allowed the web examples of the originally formulated sentence as well as the recommended correction. The error types mainly targeted were the presence and choice of article, preposition, noun number, gerund or infinitive confusion, auxiliary verbs, verb inflections, local word order and adjective or noun confusion. The system using native English corpus was evaluated automatically and manual evaluation was done on web writing corpus, email writing corpus and Chinese Learners' of English Corpus (CLEC). The accuracy of article choice classifier was 86.06% and the accuracy for the presence and absence of preposition was 84.54% and 91.81% respectively. Results were quite encouraging and differed according to the user input text which refers to the drawback of training the system on generic native data.

Foster & Andersen [38] address the issues involved in generating ungrammatical data and present an error generation tool known as GenERRate. Errors were generated by inserting a word, deleting a word, moving and substituting a word. The effects of the error data generated by the proposed tool on the performance of classifier were studied. For this purpose, two experiments were carried out, one with the original Cambridge learner corpus (CLC) and the other with data that contains artificially introduced errors. Comparison of the results indicated 6.2% drop in accuracy of the classifier by moving from training on original corpus to synthetic corpus. However, to recover the performance degradation, it is suggested to use a hybrid of the two types of corpora where synthetic ungrammatical data could be used to expand the original learner training datasets.

Rozovskaya & Roth [20] carried out research mainly focused on the correction of articles in the English text. They advise different error training models and show that their model performs better than the native data training model. Their results also show that those error correcting models operate a lot better that have information on the division of article and non-native phrases including patterns of error.

A new strategy for GEC based on alternating structure optimization (ASO) approach has been proposed by Dahlmeier & Ng [33]. ASO is a multi-task learning algorithm that makes use of the common structure of multiple related tasks. The data set used by them was NUCLE. Their results portrayed that for article and preposition error the ASO strategy performed much better than two pre-existing commercial grammar checking software.

Chodorow & Leacock [39] put forward an unsupervised method that could find grammatical errors by deducing undesirable data from the edited corpus. Their experiment was directed at the identification of irregular use of certain words of the English vocabulary in the Test of English as a Foreign Language (TOEFL) essays. The system that they developed was called ALEK that is short for Assessing Lexical Knowledge, and it made use of statistical methods for analysis. The aim was to identify extra words in the sentence, any missing word, any wrong word, form of verb, punctuation errors, spelling errors, sentence fragment errors and word form errors. The performance of the system after carrying out certain experiments was found to be at 80% precision and a recall of about 20%. Hence, they concluded that ALEK was very productive in identifying the errors in the text.

Although classification approach to GEC has proven to be quite successful; it also has some shortcomings. As this approach requires a separate classifier for each error type so, the classifier cannot apply multiple corrections simultaneously rather it can correct one word at a time for some specific error category e-g preposition, verb, noun etc. Secondly, it assumes the context of the rest of the sentence correct which is not the case in real time.

3.2.2. Statistical Machine Translation Approach

An alternative way is to consider the error correction task as a translation problem. Statistical machine translation is a machine translation approach which uses a large dataset of translations from one language to another and infers a statistical model from these sets of translations. This statistical model can then be applied to new texts to make predictions. This technique can be employed for GEC by using a corpus containing a large number of grammatically incorrect sentences already converted to multiple correct phrases. The statistical model derived from the analysis of this parallel corpus can be applied to new erroneous sentences to make a guess to a reasonable correction. For example, an erroneous text e is given and we would like to find a good grammatical correction c of the given text. There could be multiple possible corrections of an erroneous sentence. This difference could be modeled with a probability distribution Pr(c|e) over possible corrections c, given that the erroneous sentence is e. The best correction can then be chosen by choosing c which maximizes the conditional probability Pr(c|e). This approach is advantageous as it does not require any explicit encoding of the features rather it learns source to target mappings from the training data keeping the context in consideration. It provides better coverage for interacting errors as it focuses on correcting the overall sentence. However, SMT cannot include models for specific error categories and requires a sufficiently large training corpus which is expensive to produce.

The study conducted by Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto [17] analysed the effect of using large size learner corpus as training data on all types of grammatical errors in ESL. The system was based on phrase-based statistical machine translation approach and was trained on KJ corpus and Lang-8 corpus. It was found that increasing the size of the corpus improved the Phrase Based Statistical Machine Translation (PBSMT) approach. However, the extent of improvement differed according to different types of errors. F-measure of 20% and 365 was attained on KT corpus and Lang-8 corpus respectively.

Yuan & Felice [24] also used PBSMT for correcting nouns, prepositions, verb form, subject verb agreement and article errors in the learner text from the subset of NUCLE V2.3 Corpus and the Cambridge Learner Corpus1 (CLC). PBSMT was trained incrementally with combinations of training data. The evaluation indicated that the system did not achieve significantly high performance i.e. F1 score of 22%; however, factors affecting the system's performance were revealed which were annotation criteria, the size of the corpus, heterogeneity and training parameters.

A year later, Felice & Yuan [37] focused on injecting artificial errors into the training corpus NUCLE v2.3 and unlike the previous works; it derives the probabilities of error generation using linguistic information, and builds a dataset to correct larger types of errors including open class errors. Moreover, variables involved in selecting the candidate sentences have also been analysed. Best results have been reported to be produced from error distributions and POS information and using hybrid datasets improves the overall results.

A language-independent strategy based upon the context of statistical machine translation was proposed by Ehsan & Faili [28]. The strategy was used to build a proofreading system that would be able to find spelling as well as grammar mistakes. To improve the efficiency of the system, it was made a hybrid with a pre-existing rule-based checker for grammar. Experiments were carried out on two languages namely English (Penn Treebank) and Persian (Peykareh corpus). Also, two types of sets were used for the evaluation of the system, one with real erroneous sentences and the other that included sentences with automatically induced errors. An elevation of almost 24% was observed in the recall, and precision metrics applied to the results.

3.2.3. Round Trip Machine Translation

Round trip machine translation deals with translating the user's native language text into some foreign language (also called the pivot language) and then translating back the foreign language to the native language. This approach being a bilingual model tends to produce better results than the unilingual models as it has the advantage of leveraging linguistic information of both the author's first and the second languages.

In as early as 1986, Ethel Schuster [29] implemented a system for correcting verb-particle and verb-prepositional phrases using English as a second language. The repairing strategy was based on the comparison of the grammar of user's native language with the grammar of English language.

Hermet and Désilets [6] applied Round Trip Machine Translation for correcting preposition errors considering French as a second language corpus and compared the correction rate of this technique to that of the unilingual model. No significant statistical difference was found between the two; however, a hybrid of the two approaches outperformed the isolated unilingual and bilingual approaches with the accuracy of 82.1%.

In contrast to the above mentioned round trip approaches, Madnani, Tetreault, & Chodorow [30] focused on using multiple pivot languages to produce different round trip translations. The work has been extended to correcting the grammatical structure of the whole sentence. The used approach is useful as it generates alternative renderings of the source sentence where each rendering besides being grammatically correct, is likely to preserve the meaning of the original sentence. English Giga Word Corpus was used as the training corpus.

3.2.4. Generation based approach

Lee & Seneff [5] presented a generation based approach to automatically correct grammatical errors of articles, prepositions, noun number, verb aspect, mode and tense by second-language learners using flight domain corpus. At sentence level, an incorrect input is taken, from which a word lattice of possible corrections is generated. After that an n-gram language model is applied to produce N-best results which are then parsed to re rank them accordingly. 88.7% of the sentences that were parsed produced indistinguishable results from the original input.

3.2.5. Joint Inference Approach

At the sentence level, different words are grammatically dependent on each other and, for an effective sentence level correction these linguistic dependencies cannot be ignored. Keeping in view the grammatical dependency problems encountered in the systems trained on independent models for specific types of errors, Rozovskaya & Roth [18] used joint interference and joint learning approach to address the linguistic inference with grammatical properties of the sentence such as article-Noun Phrase head and subject -verb agreement. The system was trained on Google corpus with word n-gram features. Integer linear programming (ILP) has been used to model the inference. The proposed approach applies correction on different types of errors jointly in a given sentence. The joint approaches significantly improved the correction rate of the system.

Another work by Wu & Ng [32] puts forward joint inference algorithm for detecting and correcting the grammatical errors namely: articles, prepositions, punctuations, noun number and spellings. Here again, (ILP) has been used to model the inference. It incorporates both the individual error classifiers and prior knowledge on GEC. The corpus used was Web 1t 5-Gram Corpus. Experiments on HOO-2011 shared task depict that the proposed approach achieves high performance and is competitive with the state of the art systems.

3.2.6. Crowdsourcing

Bernstein, et al. [34] present a word processing interface named Soylent which provides text shortening service, human aided spelling and grammar checker and, on demand documents editing service. The main contribution of this work is to embed crowd workers in the user interface to allow handling of tasks on demand. The grammar checker service finds the errors of the selected section of text, explains the errors and provides five alternatives for rewriting. The system worked with an average accuracy of 70.8%.

Pavlick, Yan, & Callison-Burch [35] in their research suggested crowdsourcing solution to suggesting grammar corrections and discussed the challenges for assuring quality in the systems. This method follows tiny procedures on words and then stores them in graph based data structure. Training data released by CoNLL 2013 shared task was used. The accuracy is measured by detaching corrections of separate entities and measuring the arguments between workers and congregate observations on edits. By using this technique we can segregate single edit phrases to check on edit-specific observation.

3.3. Hybrid Approaches

A hybrid approach means that rather than using only a single approach, two or more approaches are used in combination to achieve better results by benefitting from each approach's advantages.

The authors of [26] combined the results of two of the most promising approaches used in GEC which are classification-based approach and SMT based approach. The grammatical errors addressed were spelling, noun number, preposition, punctuation, article, verb form, subject-verb agreement and the corpora used were NUCLE and lang-8 corpus of learner English. Using the test set of CoNLL-2014 shared task, the system yielded F0.5 score of 39.39%.

Dahlmeier & Ng [25] proposed a beam-search decoder for GEC that used a combination of the advantages of SMT approach i.e. it could focus on correcting the overall sentence by learning source to target mappings and the classification based approach that helped with the correction of article and preposition errors if any within the complete sentence. Their model could carry out end-to-end correction of entire sentences, deal with multiple inter-connected errors such as spellings, articles, prepositions, noun number and punctuation insertion, and was trained discriminatively on Web 1T 5-gram corpus. Furthermore, it integrated existent classifier-based models for error correction. The decoder obtained a 25.48% correction score which was the best result at the time in the help our own (HOO) shared task.

Xiang, Yuan, Zhang, Wang, Zheng, & Wei [31] presented a hybrid model for correcting determiners, prepositions, noun form, and verb form and subject-verb agreement. They made use of different modules that either applied rule-based or machine-learning based strategies to achieve the task at hand. The classifier was trained using maximum entropy model using NUCLE corpus. The experiments showed that when pre and post-processing procedures were applied on the phrases, the precision, recall and F1 values observed were significantly improved. The F1 score yielded by the model was 27.3%.

Rozovskaya, Roth, & Srikumar [27] propose a linguistically-motivated approach to correct verb errors made by English as second language learners using FCE corpus. Statistical Machine Learning approach has been incorporated in rule based system which encodes linguistic information. The proposed approach first determines the verb candidates in the learner text and then, working on the concept of verb finiteness, identifies and characterizes the error. It has been demonstrated that the linguistically-informed model enhances the accuracy of verb correction.

4. GEC in Spoken Language

Grammatical error detection and providing feedback have proven to be complicated issues in the written text. When we consider the task of correcting in spoken language, it becomes even more challenging because the errors induced by automatic speech recognition (ASR) also become involved including multiple false alarms as many a times an utterance that is grammatically correct may be judged as incorrect due to a recognition error rather than a grammatical mistake.

Lee, Ryu, Seo, Kim, & Lee [22] in their research presented a model for a GEC system using JLE error-tagged corpus that made use of partially observable Markov decision process (POMDP) to provide accurate feedback for a language learning system that was dialog based. The goal was to support the correction of grammatical errors such as incorrect prepositions, determiner errors, incorrect verb form and agreement. The prototype showed positive results, in that it reduced the error rate of identifying a correct spoken dialogue as incorrect.

Izumi, Uchimoto, Saiga, Supnithi, & Isahara [4] described a method to detect grammatical as well as lexical errors made by the Japanese learners of the English language in spoken data. Audio recorded data extracted from an interview test, the "Standard Speaking Test (SST)" was used as the corpus data. Methods that increased the accuracy of error detection with a certain amount of limited training data were also described. Error categories to be detected include nouns, verbs, adjectives, adverbs, prepositions, articles, pronouns and collocations. The recall rate was found at 30% and the precision rate was at approximately 50% when the corpus was used in its initial state. However, when corrected sentences and intentional errors were introduced in the corpus that was made up of sentences Extracted from Standard Speaking Test. The precision rate improved by another 30% but there was no change in the results of the recall rate.

5. Metrics used for Evaluation of Techniques

Some commonly used metrics for evaluation of GEC systems are listed in Table 5. The research works are categorized on the basis of technique used to present a general idea of use of each metric.

The metrics used for the evaluation of GEC systems mostly include recall, precision and F-measure. Precision measures the relevancy of the results while recall measures how many actual relevant results have been returned. The formulas for calculating precision and recall are given in equation (1) and equation (2) respectively.

$$\text{Precision} = \frac{\quad}{\quad + \quad} \qquad (1)$$

$$\text{Recall} = \frac{\quad}{\quad + \quad} \qquad (2)$$

Where, TP stands for true positive, FP stands for false positive and FN stands for false negative.

F-measure also called F score contains both the recall and the precision to calculate the score. It is the harmonic mean of precision and recall. The formula for calculating F-measure is given in equation (3).

$$F_1 = 2.\frac{\quad}{\quad \cdot \quad} \qquad (3)$$

The authors of [40] argued that the metrics used for evaluation of GEC systems should take into consideration factors such as data skewing and the application type that the system will be used for. They examined ways to find appropriate methods to report the results of the system's review; these recommendations rely upon making distinct one's assumptions and applications for the identification of errors. Further research and application in this area can help in catering the evolution of problems of grammatical error detection.

Dahlmeier et al., [36] presented a new algorithm named Max Match (M2) for the purpose to evaluate GEC. The algorithm aimed to effectively calculate the string of phase-level edits that was the nearest to the highest overlap with the gold-standard annotation. They carried out experiments on the data from the Helping Our Own (HOO) shared tasks and scored them by the F1 measure; their results showed that the algorithm overcame the drawbacks in the existing evaluation method.

6. Extent of Error Correction

All the employed techniques for evaluating GEC aim at the recognition and solution of some certain type of errors e.g. verb or preposition misuses [27, 23, 5, 4, 31, and 17].
The extent of error correction of these works can be examined in table 6, which shows the grammatical categories taken into account by the researchers for the development of GEC system.

Table 6 shows that most frequent grammatical categories worked on by researchers are noun, form of verb, subject-verb agreement, use and misuse of article and choice of prepositions. However, Gamon et al., [3] considered a number of other error categories as well while [35] claims to consider all types of errors.

Considering some specific categories limits the results of GEC, Park et al., [23], Dahlmeier et al, [25] and Madani et al., [30] overcame this limitation by correcting errors at whole sentence level. Park et al., [23] used expectation-maximization (EM) algorithm for parameter training. Results were analysed using Machine Translation (MT) assessment techniques.

Table 1. Common error types encountered in grammatical error correction

Error Type	**Explanation**	**Erroneous Sentence**	**Correct Sentence**
Singular and Plural Nouns Errors	This error occurs when the writer is not sure about which nouns are countable and which are uncountable.	My hairs are greying.	My hair is greying.
Subject Verb Agreement Errors	These errors occur when there is disagreement between subject and verb in number or person.	He eat an apple daily	He eats an apple daily
Verb Form Errors	Occur when the correct form of verb is not used.	Children will flew kites in the evening.	Children will fly kites in the evening.
Verb Tense Errors	Occur when time marker is incorrect.	I am playing with Bob since morning.	I have been playing with Bob since morning.

Article Errors (Presence and Choice)	Occur when the choice of article used in the sentence is incorrect or when the appropriate article is missing.	I live in the Islamabad.	I live in Islamabad.
Preposition (Presence and Choice)	Occur when the preposition is absent or an incorrect preposition is used.	The frog jumped on the water.	The frog jumped into the water.

Table 2. Distribution of Errors in KJ Corpus

Error Types	**Proportion (%)**	**Types**	**Proportion (%)**
Article	19.23	Verb other	4.09
Noun number	13.88	Adverb	3.59
Preposition	13.56	Conjunction	2.04
Tense	8.77	Word order	1.34
Lexical choice of noun	7.04	Noun other	1.30
Lexical choice of verb	6.90	Auxiliary verb	0.88
Pronoun	6.62	Other lexical choice	0.74
Agreement	5.25	Relative	0.42
Adjective	4.30	Interrogative	0.04

Table 3: Distribution of Errors in CoNLL-2013 Corpus

Error Types	**Error Rates (%)**	
	Training	**Testing**
Article	2.4	10.0
Preposition	2.0	10.7
Noun	1.6	6.0
Verb Agreement	2.0	5.2
Verb Form	0.8	2.5

Table 4: Techniques used in the research works, techniques have not been grouped under the main category and have been listed specifying the sub technique used.

Technique	**Research Works**

Rule Based Approach	Sidorov, Gupta, Tozer, Catala, Catena, & Fuentes [8], Nagata, Morihiro, Kawai, & Isu [7], Sidorov, Grigori [19]
Classification	Rozovskaya & Roth [20]
Classification Using Maximum Entropy Model	Rozovskaya & Roth [21], Tetreault & Chodorow [12], Izumi, Uchimoto, Saiga, Supnithi, & Isahara [4]
Decision Tree	Gamon, et al. [10]
Partially Observable Markov Decision Process	Lee, Ryu, Seo, Kim, & Lee [22]
Noisy Channel Models	Park & Levy [23]
Generation-Based Approach	Lee & Seneff [5]
Statistical Machine Translation (SMT)	Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto [17] , Yuan & Felice [24]
Hybrid Of Classification And SMT	Dahlmeier & Ng [25], Susanto [26]
Rule Based System And SMT	Rozovskaya, Roth, & Srikumar, 2014 [27], Ehsan & Faili [28]
Round Trip Machine Translation	Schuster [29], Hermet & D'esilets [6], Madnani, Tetreault, & Chodorow [30]
Hybrid Of Rules Based Approach And Machine Learning	Xiang, Yuan, Zhang, Wang, Zheng, & Wei [31]
Joint Inference	Wu & Ng [32]
Alternating Structure Optimization (ASO)	Dahlmeier & Ng [33]

Joint Inference And Joint Learning	Rozovskaya & Roth [18]
Multi-Outcome Prediction	Han, Tetreault, Lee, & Ha [11]
Microsoft Research (MSR) ESL	Gamon, et al. [3]
Crowd Sourcing	Bernstein, et al. [34], Pavlick, Yan, & Callison-Burch [35]
Mass Count Distinction	Nagata, Morihiro, Kawai, & Isu [7]
M2 Algorithm	Dahlmeier & Ng [36]
Injection Of Artificial Errors Using Conditional Probabilities	Felice & Yuan [37]

Table 5: Evaluation Metrics used for evaluation of GEC systems

Evaluation Metrics	**Research Works**
Accuracy	Bernstein, et al. [34]
F1 Measure	Dahlmeier & Ng [36]
Recall	Rozovskaya, Roth, & Srikumar [27], Gamon, et al. [3], Lee, Ryu, Seo, Kim, & Lee [22]
Repair Rate	Hermet & D'esilets [6]
Precision And Recall	Gamon, et al. [10], Felice & Yuan [37], Chodorow & Leacock [39], Nagata, Morihiro, Kawai, & Isu [7], Rozovskaya & Roth [20], Han, Tetreault, Lee, & Ha [11], Tetreault & Chodorow [12], Ehsan & Faili [28], Izumi, Uchimoto, Saiga, Supnithi, & Isahara [4]
Meteor Score, Bleu Score	Park & Levy [23]

Performance, Reliability	Pavlick, Yan, & Callison-Burch [35]
	Dahlmeier & Ng [25], Sidorov, Gupta, Tozer, Catala, Catena, & Fuentes [8],
Precision, Recall, F1 Measure	Sidorov, Grigori [19], Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto [17], Xiang, Yuan, Zhang, Wang, Zheng, & Wei [31], Wu & Ng [32], Jia, Wang, & Zhao [21], Dahlmeier & Ng [33], Yuan & Felice [24]
Accuracy, Precision, Recall, F-Score	Lee & Seneff [5]
Success Rate, Draw Rate, Failure Rate	Madnani, Tetreault, & Chodorow [30]
Accuracy, Precision, Recall, Performance	Rozovskaya & Roth [18]
Accuracy, Precision, Recall, F-Measure, Meteor	Susanto [26]

Table 6: Extent of error correction employed by research works

Papers	**Extent Of Error Correction**
Rozovskaya & Roth [18]	Article, Noun, Phrase head and subject-verb
Lee, Ryu, Seo, Kim, & Lee [22]	Prepositions, determiners, verb form and agreement
Park & Levy [23]	Spellings, article, prepositions, and insertion errors + whole Sentence
Lee & Seneff [5]	Articles, prepositions, noun number, verb aspect, mode and tense
Dahlmeier & Ng [25]	Spellings, articles, prepositions, noun number and punctuation insertion+ whole Sentence

Izumi, Uchimoto, Saiga, Supnithi, & Isahara [4]	Nouns, verbs, adjectives, adverbs, prepositions, articles, pronouns and collocations
Xiang, Yuan, Zhang, Wang, Zheng, & Wei [31]	Determiners, prepositions, noun form, verb form and subject-verb agreement
Wu & Ng [32]	Articles, prepositions, punctuations, noun number and spellings
Jia, Wang, & Zhao [21]	Determiners, prepositions, modal verbs, noun number, verb number, verb tense
Susanto [26]	Spelling, noun number, preposition, punctuation, article, verb form, subject-verb agreement
Gamon, et al. [3]	Article, preposition, noun number, gerund or infinitive confusion, auxiliary verbs, verb inflections, local word order and adjective or noun confusion
Chodorow & Leacock [39]	Missing word, wrong word, form of verb, punctuation errors, spelling errors, sentence
Yuan & Felice [24]	Nouns, prepositions, verb form, subject verb agreement and article errors
Felice & Yuan [37]	Nouns, prepositions, verb form, subject verb agreement, article errors and open class errors
Ehsan & Faili [28]	Prepositions, conjunctions, noun count, verbs, adjectives, adverbs, subject-verb agreement, repeated words, context sensitive spellings
Sidorov, Gupta, Tozer, Catala, Catena, & Fuentes [8]	Noun number, subject-verb agreement, verb form, article/determiner and preposition
Sidorov, Grigori [19]	Noun number, subject-verb agreement, verb form, article/determiner and preposition
Madnani, Tetreault, & Chodorow [30]	Whole Sentence

Nagata, Morihiro, Kawai, & Isu [7]	Mass and count nouns
Dahlmeier & Ng [36]	Phrase level correction
Gamon, et al. [10]	Determiner, preposition
Dahlmeier & Ng [33]	Articles and prepositions
Ethel Schuster [29]	Verb-particle and verb-prepositional phrases
Bernstein, et al. [34]	Spellings + other errors
Rozovskaya & Roth [20]	Articles
Tetreault & Chodorow [12], Han, Tetreault, Lee, & Ha [11], Hermet & D'esilets [6]	Prepositions
Mizumoto, Hayashibe, Komachi, Nagata, & Matsumoto [17], Pavlick, Yan, & Callison-Burch [35]	All types of errors
Rozovskaya, Roth, & Srikumar [27]	Verb

Conclusions

The use of automatic tools for the recognition as well as the solution of grammatical mistakes made by language learners has escalated over the past years. This escalation can be largely attributed to the fact that the number of foreign language learners has grown considerably. According to an estimate, more than a billion people use English as a foreign language. These statistics push the natural language processing (NLP) community towards developing automated applications that can help second language learners to write as well as speak efficiently in the non-native language.

Over the years, many different methods and techniques have been researched with the goal to build a system that maximizes the benefits and diminishes the disadvantages associated with the various techniques such as round trip machine translation [24-26], classification [12-14], rule based approach [8, 10], and statistical machine translation [6, 19].

The paper provides a detailed review and analysis of these various techniques and methods that are used for automatic GEC. It studies the research carried out over the years by many researchers that have employed these different techniques to make GEC systems and evaluating them by using data sets from various corpora such as NUCLE, CLC etc. to get an idea of the performance of the system. Starting with category wise brief summary of the work, the papers have been categorized on the basis of techniques used, evaluation metric used and extent of error correction.

From the review it can be concluded that among all these techniques, round trip machine translation technique [24-26] offers bilingual analysis compared to all the other techniques that only work on one language. The concept behind this approach is that it takes into account the structure of both the native and the second language of the learner to ensure that the writer is not making mistakes in his native language. Hence, it can be said that this technique is more effective in this regard.

References

[1] J. S. Y. Lee, "Automatic correction of grammatical errors in non-native English text," 2009.

[2] G. Heidorn, "Intelligent writing assistance" Handbook of natural language processing, pp. 181-207, 2000.

[3] M. Gamon, C. Leacock, C. Brockett, W. B. Dolan, J. Gao, D. Belenko and A. Klementiev, "Using statistical techniques and web search to correct ESL errors,"

Calico Journal, vol. 26, no. 3, pp. 491-511, 2013.

[4] E. Izumi, K. Uchimoto, T. Saiga, T. Supnithi and H. Isahara, "Automatic error detection in the Japanese learners' English spoken data," in Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 2, 2003.

[5] J. Lee and S. Seneff, "Automatic grammar correction for second-language learners.," in INTERSPEECH, 2006.

[6] M. Hermet and A. D'esilets, "Using first and second language models to correct preposition errors in second language authoring," in Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications, 2009.

[7] R. Nagata, K. Morihiro, A. Kawai and N. Isu, "A feedback-augmented method for detecting errors in the writing of learners of English," in Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, 2006.

[8] G. Sidorov, A. Gupta, M. Tozer, D. Catala, A. Catena and S. Fuentes, "Rule-based System for Automatic Grammar Correction Using Syntactic N-grams for English Language Learning (L2)," CoNLL-2013, p. 96, 2013.

[9] P. S. Gingrich, "Human factors and behavioral science: The UNIX™ Writer's Workbench software: Results of a field study," Bell System Technical Journal, The, vol. 62, no. 6, pp. 1909-1921, 1983.

[10] M. Gamon, J. Gao, C. Brockett, A. Klementiev, W. B. Dolan, D. Belenko and L. Vanderwende, "Using Contextual Speller Techniques and Language Modeling for ESL Error Correction.," in IJCNLP, 2008.

[11] N.-R. Han, J. R. Tetreault, S.-H. Lee and J.-Y. Ha, "Using an Error-Annotated Learner Corpus to Develop an ESL/EFL Error Correction System.," in LREC, 2010.

[12] J. R. Tetreault and M. Chodorow, "The ups and downs of preposition error detection in ESL writing," in Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1, 2008.

[13] R. Dale and A. Kilgarriff, "Helping our own: The HOO 2011 pilot shared task," in Proceedings of the 13th European Workshop on Natural Language Generation, 2011.

[14] R. Dale, I. Anisimoff and G. Narroway, "HOO 2012: A report on the preposition and determiner error correction shared task," in Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, 2012.

[15] H. T. Ng, S. M. Wu, Y. Wu, C. Hadiwinoto and J. Tetreault, "The conll-2013 shared task on grammatical error correction," in Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task (CoNLL-2014 Shared Task), 2014.

[16] S. D. Hernandez and H. Calvo, "CoNLL 2014 shared task: Grammatical error correction with a syntactic n-gram language model from a big corpora," CoNLL-2014, p. 53, 2014.

[17] T. Mizumoto, Y. Hayashibe, M. Komachi, M. Nagata and Y. Matsumoto, "The effect of learner corpus size in grammatical error correction of ESL writings," 2012.

[18] A. Rozovskaya and D. Roth, "Joint learning and inference for grammatical error correction," Urbana, vol. 51, p. 61801, 2013.

[19] Sidorov, Grigori, "Syntactic dependency based n-grams in rule based automatic English as second language grammar correction," International Journal of Computational Linguistics and Applications, vol. 4, no. 2, pp. 169-188, 2013.

[20] A. Rozovskaya and D. Roth, "Training paradigms for correcting errors in grammar and usage," in Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics, 2010.

[21] Z. Jia, P. Wang and H. Zhao, "Grammatical Error Correction as Multiclass Classification with Single Model," CoNLL-2013, p. 74, 2013.

[22] K. Lee, S. Ryu, P. H. Seo, S. Kim and G. G. Lee, "Grammatical error correction based on learner comprehension model in oral conversation," in Spoken Language Technology Workshop (SLT), 2014 IEEE, 2014.

[23] Y. A. Park and R. Levy, "Automated whole sentence grammar correction using a noisy channel model," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011.

[24] Z. Yuan and M. Felice, "Constrained grammatical error correction using Statistical Machine Translation," CoNLL-2013, p. 52, 2013.

[25] D. Dahlmeier and H. T. Ng, "A beam-search decoder for grammatical error correction," in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and

Computational Natural Language Learning, 2012.
[26] [26] R. H. Susanto, "Systems Combination for Grammatical Error Correction," 2015.
[27] [27] A. Rozovskaya, D. Roth and V. Srikumar, "Correcting grammatical verb errors," in Proceedings of EACL, 2014.
[28] N. Ehsan and H. Faili, "Grammatical and context-sensitive error correction using a statistical machine translation framework," Software: Practice and Experience, vol. 43, no. 2, pp. 187-206, 2013.
[29] E. Schuster, "The role of native grammars in correcting errors in second language learning," Computational Intelligence, vol. 2, no. 1, pp. 93-98, 1986.
[30] N. Madnani, J. Tetreault and M. Chodorow, "Exploring grammatical error correction with not-so-crummy machine translation," in Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, 2012.
[31] Y. Xiang, B. Yuan, Y. Zhang, X. Wang, W. Zheng and C. Wei, "A hybrid model for grammatical error correction," in Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task, 2013.
[32] Y. Wu and H. T. Ng, "Grammatical Error Correction Using Integer Linear Programming.," in ACL (1), 2013.
[33] D. Dahlmeier and H. T. Ng, "Grammatical error correction with alternating structure optimization," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011.
[34] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell and K. Panovich, "Soylent: a word processor with a crowd inside," Communications of the ACM, vol. 58, no. 8, pp. 85-94, 2015.
[35] E. Pavlick, R. Yan and C. Callison-Burch, "Crowdsourcing for grammatical error correction," in Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work \& social computing, 2014.
[36] D. Dahlmeier and H. T. Ng, "Better evaluation for grammatical error correction," in Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2012.
[37] M. Felice and Z. Yuan, "Generating artificial errors for grammatical error correction," EACL 2014, p. 116, 2014.
[38] J. Foster and {. E. Andersen, "GenERRate: generating errors for use in grammatical error detection," in Proceedings of the fourth workshop on innovative use of nlp for building educational applications, 2009.
[39] M. Chodorow and C. Leacock, "An unsupervised method for detecting grammatical errors," in Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, 2000.
[40] M. Chodorow, M. Dickinson, R. Israel and J. R.Tetreault, "Problems in Evaluating Grammatical Error Detection Systems," in COLING, 2012