IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 1, January 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

Preposition Error Correction Using Iterative Sequence Tagging

Shikha Prajapati, Ashish Kumar Tiwari, Madhvi Soni, Jitendra Singh Thakur

Department of Computer Science and Engineering, Jabalpur Engineering College, Jabalpur, India

ABSTRACT: Prepositions are one of the most challenging units encountered by second language learners of English. They are frequently used and are highly polysemous in nature. Moreover, their usage is highly dependent upon the underlying context of the sentence and the intent of the speaker which turns out to be a major challenge in developing a preposition correction system. In this paper, we present a preposition error correction system which is a sequence tagger based on a transformer encoder that predicts token-level edit operations to remove preposition errors from a given sentence. Our system is trained over synthetic data constructed using Google's One-billion-word benchmark corpus. For training, we used a new optimiser - Ranger which is a combination of Rectified Adam and Look Ahead optimiser. It helped to reduce the time required for training by about 6 folds. Our system outperformed the state-of-the-art approaches and achieved precision, recall and $f_{0.5}$ score of **91.4, 77.27 and 88.27** respectively on a dataset prepared from preposition exercises collected from different websites and grammar books.

I. INTRODUCTION

Prepositions are words that show some connection between different parts of a sentence. They usually come after nouns, (or pronouns, noun phrases, gerunds or noun clauses beginning with a wh-word or how), adjectives and some verbs [53]. Although the standard position for a preposition is immediately before its object, for instance : “He put the box on the shelf.” However, in some instances, a prepositional phrase, which is a preposition plus its object, can also begin sentences [19], for instance : “After January comes February.”

Determining correct preposition usage is inherently challenging due to their diverse linguistic functions. Indeed, [20] found only ~75% agreement among native speakers and reference texts on fill-in-the-blank tasks using Encarta Encyclopedia sentences [25], highlighting the task's complexity for learners. Consequently, preposition errors form a major portion of learner mistakes: studies show they account for 29% of all errors made by English learners [21], appear in 8% of analyzed sentences [22], and occur at a ~10% rate in Japanese learner corpora [23]. Consequently, because preposition errors are the most common mistake among learners, many automated grammar checking tools are specifically designed to detect and correct them [54], [55].

Preposition errors the one of the most common types of errors made by second languagecan be broadly classified into three types namely, i) extraneous error, where a preposition has been extraneously inserted, ii) missing error, where an essential preposition has been skipped and iii) replacement error, where a preposition has been wrongly selected.

Over the years, various statistical preposition error correction methods have relied on native English corpora or hand-annotated learner data [1]–[3], [20], [23], [25], [26], [29], [31], [34]–[36], though building such annotated datasets is time-consuming and labor-intensive. Recently, Neural Machine Translation [5] and Transformer-based sequence-to-sequence models [6] have yielded state-of-the-art results on general grammatical error correction benchmarks [6]–[9], [11]–[13]. However, few studies have specifically evaluated these architectures on preposition errors alone, and their practical adoption remains hindered by slow training speeds and heavy reliance on massive annotated datasets.

In this paper, we have used an iterative sequence tagger [13] based on a transformer encoder to predict token-level edit operations [4] for correcting the preposition errors in a sentence. For each source token in the input sentence, a softmax layer [4] on the top of the encoder predicts an edit operation out of a confusion set. The system predicts four basic edit operations namely keep a token, delete a token, append a token and replace a token. The system is trained over a synthetic dataset constructed by us using Google's one billion word corpus[37]. Additionally, we have incorporated in our training procedure the new state-of-the-art Ranger optimiser which is a combination of Rectified Adam and lookahead [38] optimiser. It is known to outperform all other optimisers for the image classification task on Imagewoof dataset. However, not much analysis has been done on the performance of the optimiser in the field of natural language processing which proved to be a motivation for us to incorporate it into our system. Some improvements of our system

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

over the previously discussed approaches are:

- Ranger optimiser helps to increase the speed of training and also the accuracy of the system.
- Synthetic data eliminates the need for large quantities of human-annotated data. Small and diverse datasets will serve the purpose of augmenting large datasets for training.

The rest of the paper is organized as follows: Section 2 presents a brief overview of the related work. Our proposed method is described in Section 3. Evaluation and experimental results are discussed in Section 4. A brief discussion over the results is given in section 5. Finally, we presented our conclusion in Section 6.

II. RELATED WORK

Prepositions comprise a large portion of verbal communication and are highly polysemous in their nature. The interaction of the above two reasons makes preposition errors the largest category of grammatical errors made by non-native speakers (accounting for nearly 29% [21]). Over the years, several approaches have been proposed for the task of preposition error correction. In 2008, [20] and [33] trained a Maximum Entropy (ME) classifier on the preposition contexts extracted from native English corpora. The features were extracted from a window of n-words before and after the prepositional part. Preposition prediction was treated as a classification task of choosing the most suitable preposition out of a fixed candidate set. Whereas [34] used a decision tree approach and a language model trained over Gigaword corpus to correct the preposition errors.

Subsequent efforts introduced deeper semantics, parsing, and learner-specific features: DAPPER [29] combined semantic, syntactic, and lexical features using a Maximum Entropy (ME) model trained on the BNC corpus [39]; [25] trained an ME classifier on a Korean learner corpus; [28] added Stanford parse features to an existing lexical system [20]; and [38] constrained candidate sets based on annotated valid corrections. Later models shifted toward diverse data sources and feedback generation: [1] applied multinomial logistic regression to Wikipedia revision edits for a 36-preposition set (limited to replacement errors), [2] used error case frames to explain faulty noun-preposition usage, and Wordprep [3] leveraged *n*-gram data mining to predict missing prepositions, though it struggles with unseen contexts. Recently, Neural Machine Translation (NMT) approaches [5]–[11], [13] have framed general grammatical error correction as translating errorneous source sentences to clean target sentences, achieving strong benchmark results [12] despite suffering from slow training times and heavy reliance on large annotated datasets. Our approach tries to overcome these problems by:

1. Constructing a large synthetic dataset from small and high-quality public datasets to train the model.
2. Incorporating the new Ranger optimiser which helps to reduce the training time and increase the accuracy of the system.

III. METHOD

Inspired by GECToR [4], our proposed system is an iterative sequence tagger [13] that uses a XLNet [40] encoder stacked with two linear layers and softmax layers. The model predicts token-level transformations $T(x_i)$ for each token x_i in an input sentence $(x_1 \dots x_N)$, which are then applied to produce the corrected text over one or more iterations to handle multiple preposition errors [4], [13]. The pipeline comprises four main steps: (1) dataset construction, (2) preprocessing, (3) training and evaluation, and (4) testing.

3.1 Construction of dataset

Although most Neural Machine Translation approaches [4], [11], [13], [41]–[43], [46], [47] train on gold-standard public datasets, these were unsuitable for our model because they contain all grammatical error types rather than exclusively preposition errors, and are too small for our training needs. Consequently, we constructed a large synthetic dataset by extracting common preposition errors and their frequencies from public corpora and injecting them into a clean, native dataset.

3.1.1 Extraction of error probabilities from public datasets

To extract common preposition errors, we used m2-formatted, automatically annotated [14] versions of public learner datasets: FCE [16], Lang-8 [15], Cambridge Learner Corpus [17], and W&I+LOCNESS [12] (featured in the BEA 2019 Shared Task [12]). From these, we extracted 112,603 sentences containing preposition errors. Ignoring other grammatical mistakes, we catalogued each preposition error, classified its type, and tracked its frequency in a dictionary. There are three main classes of errors (refer Table 1):

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

- Extraneous: An extraneous preposition has been used where it is not required.
- Missing: An essential preposition has been omitted.
- Replacement: A preposition has been wrongly replaced with another preposition.

Table 1: Distribution Of Errors

Type of error	No. of errors in each dataset		Percentage out of 100
	Public	Synthetic	
Extraneous	53721	1169458	58%
Missing	101100	544402	27%
Replacement	215011	302447	15%
Total	369832	2016307	100%

3.1.2 Introduction of errors in a clean dataset

After extracting types and frequencies of the commonly occurring preposition errors, we introduced them into a clean native dataset extracted from One billion Word benchmark [36] by Google using the algorithm given below. Consider the clean dataset to be denoted by T. E is a dictionary containing probabilities of errors according to class type. C is a dictionary containing probabilities of errors according to their count in a sentence.

For each sentence t_i in the clean dataset T : input sentence = t_i
Select the error count (c) from C with probability P(c)
Execute c times
Select the error type e from E with probability P(e) output sentence = errorify(input sentence, e) input sentence = output sentence
Write incorrect and correct sentences in separate files

3.2 *Preprocessing*

After constructing parallel datasets of incorrect and correct sentences, we need to convert them into tags so that our model can be trained for sequence tagging[4]. The parallel files of incorrect and correct sentences are converted into tags through the following procedure : for each source token x_i from the source sentence ($x_1 \ldots x_N$), $1 \leq i \leq N$, we search for the best-fitting subsequence $Y_i = (y_{j1} \ldots . y_{j2})$, $1 \leq j1 \leq j2 \leq M$ of the target sentence ($y_1 \ldots y_M$) by minimizing the modified Levenshtein distance. Each mapping is assigned a token- level transformation which when applied to the source token yields the subsequence of the target sentence. We have used the basic transformations formed by [4]. These include (tag$KEEP) keep the current token unchanged, (tag$DELETE) delete the current token, (tag$APPEND_t1) append new token t1 next to the current token xi or (tag$REPLACE_t2) replace the current token xi with another token t2 [4]. An example of a mapping from source to target sentence is given below :

Source sentence: The book is kept in the table. Target sentence: The book is kept on the table.
Mapping: [The →The], [book → book], [is→ is], [kept→ kept], [in → on], [the → the], [table → table]
Transformations list:

- [The →The], [book → book], [is→ is], [kept→ kept], [the → the], [table → table] : $KEEP,
- [in → on] : $REPLACE_on

These transformations are fed to the encoder model through which it learns to predict the transformations required for each token in a given incorrect sentence.

3.3 Training

After preprocessing the data we feed it to our encoder model for fine- tuning. We used a pre- trained transformer encoder (xlnet [40]) in its base configuration with default hyper parameters to train for the task of sequence tagging. We used Ranger optimiser with gradient centralization applied to all the layers. The data was split into train and dev sets with 98:2 ratio respectively. Training was performed for 5 epochs with a batch size of 64. We fine-tuned our transformer models in Google COLAB.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

IV. RESULTS

We evaluated our models on the CoNLL- 2014 test set, BEA- 2019 test dataset and on a dataset of 106 sentences containing preposition exercises collected from the web and grammar books. We evaluated our system with ERRANT[14]. A comparison of our results with other gec tools is given in Table 2. It is clear from the results that our model performs better than other approaches on the third dataset containing preposition exercises. It is also noticeable that the f-score is comparatively less than other approaches when tested over the public datasets [12].

Table 2. Comparison of our results with other gec systems.

Preposition Correction Tool	BEA -2019 test set[a]			CoNLL-2014 test set			Our test dataset		
	P	**R**	**$F_{0.5}$**	**P**	**R**	**$F_{0.5}$**	**P**	**R**	**$F_{0.5}$**
GECToR (xlnet)	64.98	71.15	66.12	46.27	33.23	42.9	81.52	68.18	78.45
Our tool	38.05	70.52	41.91	50.66	11.48	30.1	91.4	77.27	88.17
Trinka	37.59	14.84	28.77	31.87	11.92	23.88	63.16	21.82	45.8
Ms - Word	75	0.97	4.6	4.37	3.13	4.05	25	1.82	7.04
Google- Docs	76.66	57.86	71.98	55.69	26.97	45.92	83.33	59.09	77.01

P: Precision, **R**: Recall , **F0.5** : F-score

[a] Codalab platform was used for evaluation on BEA-2019 test set and span-level correction score corresponding to the preposition error type was considered because the annotated file is not available publicly.

V. DISCUSSION

Our system clearly outperforms existing models on web-collected preposition exercises containing only preposition errors. However, its performance drops on public benchmarks because extracted sentences often contain non-prepositional errors. These extrinsic errors can alter sentence meaning or overlap with prepositional contexts, leading to incorrect predictions. For example:

Incorrect: *I started invoving into Facebook one years ago, when I just arrived singapore.*
Correct: *I started becoming involved in Facebook one year ago, when I just arrived in Singapore.*
Our System's Output: *I started invoving into Facebook one years ago, when I just arrived in singapore.*
There are other instances in the test dataset where a preposition/prepositional phrase needs to be replaced by a non-prepositional token but the model tries to correct it by replacing it with another preposition. For example:

Incorrect: *If certain disease genetic test is very accurate and it is unavoidable and necessary to get treatment and known by others , it is OK to disclose the result .*
Correct: *If a certain genetic test is very accurate and it is unavoidable and necessary to get treatment and tell others , it is OK to disclose the result*
Our System's Output: *If certain disease genetic test is very accurate and it is unavoidable and necessary to get treatment and known to others , it is OK to disclose the result .*

There are other instances where an entire phrase in the incorrect sentence needs to be replaced with another phrase containing prepositions and verbs. For example:

Incorrect: *During that period , if one of the family member reflects genetic disorder symptoms , he will fell in an ethical dilemma for sure .*
Correct: *During that period , if one of the family members was affected by genetic disorder symptoms, he faced an ethical dilemma .*
Our System's Output: *During that period , if one of the family member reflects genetic disorder symptoms , he will fell into an ethical dilemma for sure .*

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

Above examples shows that in practical situations preposition errors cannot exist in isolation and a preposition correction system needs to be integrated with a bigger grammatical error correction system to suggest corrections for all the above-mentioned scenarios. However, our system performs comparatively better than other approaches for sentences having only preposition errors. Therefore, it can be useful for providing learners with preposition exercises to learn its usage. Our system can also be integrated with a bigger grammatical error correction system to improve its performance.

Meanwhile, the motivation behind using Ranger optimiser was that it is a recent optimiser and has achieved state-of-the-art results for image classification tasks. However, less work has been done to test the performance of the optimiser in the field of natural language processing. The optimiser is a combination of Rectified Adam and Look Ahead [38] optimiser with gradient centralization [52]. We found that Ranger helped to decrease the time required for training considerably. It took 3 hours for training with Adam for a single epoch whereas it took only 2.5 hours for training for 5 epochs with Ranger. Ranger also helped to achieve state-of-the-art results on the constructed preposition dataset.

VI. CONCLUSIONS AND FUTURE WORK

We used a pretrained transformer encoder with iterative sequence tagging for preposition error correction and achieved state-of-the-art results on a dataset of preposition exercises collected from the web with a precision, recall and an $F_{0.5}$ score of 91.4, 77.27 and 88.17 respectively. We also showed that preposition errors do not exist in isolation by evaluating our model over public benchmark datasets and hypothesized that a preposition correction model must be integrated with a grammatical error correction model to cover a broader spectrum of preposition errors. We showed that the Ranger optimiser can be used in the natural language processing tasks for faster training. In our case it helped to decrease the training time by 6 folds. We encourage our system to be used in teaching systems for demonstrating preposition usage to the learners. Our system can also be integrated with a grammatical error correction system.

REFERENCES

[1] Cahill, A. et al. “Robust Systems for Preposition Error Correction Using Wikipedia Revisions.” HLT-NAACL (2013).
[2] Ryo NAGATA, Edward WHITTAKER, A Method for Correcting Preposition Errors in Learner English with Feedback Messages, IEICE Transactions on Information and Systems, 2017.
[3] P. Bhagat, A. S. Varde and A. Feldman, "WordPrep: Word-based Preposition Prediction Tool," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 2169-2176, doi: 10.1109/BigData47090.2019.9005608.
[4] Omelianchuk, Kostiantyn & Atrasevych, Vitaliy & Chernodub, Artem & Skurzhanskyi, Oleksandr. (2020). GECToR – Grammatical Error Correction: Tag, Not Rewrite. 163-170. 10.18653/v1/2020.bea-1.16.
[5] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Edinburgh neural machine translation systems for WMT 16. In Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers, pages 371–376, Berlin, Germany. Association for Computational Linguistics.
[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008
[7] Wei Zhao, Liang Wang, Kewei Shen, Ruoyu Jia, and Jingming Liu. 2019. Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data. arXiv preprint arXiv:1903.00138.
[8] Jared Lichtarge, Chris Alberti, Shankar Kumar, Noam Shazeer, Niki Parmar, and Simon Tong. 2019. Corpora generation for grammatical error correction. arXiv preprint arXiv:1904.05780.
[9] Tao Ge, Furu Wei, and Ming Zhou. 2018b. Reaching human-level performance in automatic grammatical error correction: An empirical study. arXiv preprint arXiv:1807.01270.
[10] Shamil Chollampatt and Hwee Tou Ng. 2018b. Neural quality estimation of grammatical error correction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2528–2539.
[11] Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and Kenneth Heafield. 2018. Approaching neural grammatical error correction as a low-resource machine translation task. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers).
[12] Christopher Bryant, Mariano Felice, Øistein E. Andersen, and Ted Briscoe. 2019. The BEA-2019 shared task on grammatical error correction. In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 52–75, Florence, Italy. Association for Computational Linguistics.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

[13] Awasthi, Abhijeet & Sarawagi, Sunita & Goyal, Rasna & Ghosh, Sabyasachi & Piratla, Vihari. (2019). Parallel Iterative
[14] Edit Models for Local Sequence Transduction. 4251-4261. 10.18653/v1/D19-1435.
[15] Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic annotation and evaluation of error types for grammatical error correction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 793–805, Vancouver, Canada. Association for Computational Linguistics.
[16] Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. 2012. Tense and aspect error correction for esl learners using global context. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short
Papers-Volume 2, pages 198–202. Association for Computational Linguistics.
[17] Helen Yannakoudakis, Ted Briscoe, and Ben Medlock. 2011. A new dataset and method for automatically grading esl texts. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 180–189. Association for Computational Linguistics.
[18] Diane Nicholls. 2003. The Cambridge learner corpus: Error coding and analysis for lexicography and elt. In Proceedings of the Corpus Linguistics 2003 conference, volume 16, pages 572–581.
[19] Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The conll-2014 shared task on grammatical error correction. In Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task, pages 1–14.
[20] Morgan, Gareth. "The effect of overt prepositional input on students' written accuracy." Advances in Language and Literary Studies 5.5 (2014): 202-212.
[21] Tetreault, Joel, and Martin Chodorow. "The ups and downs of preposition error detection in ESL writing." Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008). 2008.
[22] Bitchener, John, Stuart Young, and Denise Cameron. "The effect of different types of corrective feedback on ESL student writing." Journal of second language writing 14.3 (2005): 191-205.
[23] Chodorow, Martin, Michael Gamon, and Joel Tetreault. "The utility of article and preposition error correction systems for English language learners: Feedback and assessment." Language Testing 27.3 (2010): 419-436.
[24] Izumi, Emi, et al. "Automatic error detection in the Japanese learners' English spoken data." The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics. 2003.
[25] Rupp-Serrano, Karen. "Microsoft Encarta." Information Technology and Libraries 13.1 (1994): 73-76.
[26] Han, Na-Rae, et al. "Using an Error- Annotated Learner Corpus to Develop an ESL/EFL Error Correction System." LREC. 2010.
[27] A. Islam and D. Inkpen, "An unsupervised approach to preposition error correction," Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering(NLPKE-2010), 2010, pp. 1-4, doi: 10.1109/NLPKE.2010.5587782.
[28] Dale, Robert, Ilya Anisimoff, and George Narroway. "HOO 2012: A report on the preposition and determiner error correction shared task." Proceedings of the Seventh Workshop on Building Educational Applications Using NLP. 2012.
[29] Tetreault, Joel, Jennifer Foster, and Martin Chodorow. "Using parse features for preposition selection and error detection." Proceedings of the acl 2010 conference short papers. 2010.
[30] Felice, Rachele De. Automatic Error Detection in Non-Native English. Oxford University, UK, 2008.
[31] Felice, Rachele De, and Stephen Pulman. "Automatic Detection of Preposition Errors in Learner Writing." CALICO Journal, vol. 26, no. 3, 2009, pp. 512–528. JSTOR, www.jstor.org/stable/calicojournal.26.3.512. Accessed 29 June 2021.
[32] Barak, Libby, et al. "Modeling Second Language Preposition Learning." (2020).
[33] Huang, Guimin, et al. "BERT-based Contextual Semantic analysis for English Preposition Error Correction." Journal of Physics: Conference Series. Vol. 1693. No. 1. IOP Publishing, 2020.
[34] De Felice, Rachele, and Stephen Pulman. "A classifier-based approach to preposition and determiner error correction in L2 English." Proceedings of the 22nd international conference on computational linguistics (Coling 2008). 2008.
[35] Michael Gamon, Jianfeng Gao, Chris Brockett, Alexander Klementiev, William Dolan, Dmitriy Belenko, and Lucy Vanderwende. 2008. Using contextual speller techniques and language modelling for ESL error correction. In Proceedings of The Third International Joint Conference on Natural Language Processing (IJCNLP 2008).
[36] Na-Rae Han, Martin Chodorow, and Claudia Leacock. 2006. Detecting errors in English article usage by non-native speakers. Natural Language Engineering, 12(2):115–129.
[37] Chelba, Ciprian, et al. "One billion word benchmark for measuring progress in statistical language modelling." arXiv preprint arXiv:1312.3005 (2013).
[38] Zhang, Michael R., et al. "Lookahead optimizer: k steps forward, 1 step back." arXiv preprint arXiv:1907.08610

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001049|

(2019).
[39] Rozovskaya, Alla, and Dan Roth. "Generating confusion sets for context-sensitive err Tetreault, Joel, and Martin Chodorow. "The ups and downs of preposition error detection in ESL writing." Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008). 2008.volume 1, pages 595–606.or correction." Proceedings of the 2010 conference on empirical methods in natural language processing. 2010.
[40] Lou Burnard, editor. 2000. The British National Corpus Users Reference Guide. British National Corpus Consortium, Oxford University Computing Services.
[41] Yang, Zhilin, et al. "Xlnet: Generalized autoregressive pretraining for language understanding." Advances in neural information processing systems 32 (2019).
[42] Kaneko, Masahiro, et al. "Encoder-decoder models can benefit from pre-trained masked language models in grammatical error correction." arXiv preprint arXiv:2005.00987 (2020).
[43] Kiyono, Shun, et al. "An empirical study of incorporating pseudo data into grammatical error correction." arXiv preprint arXiv:1909.00502 (2019).
[44] Grundkiewicz, Roman, and Marcin Junczys-Dowmunt. "Near human-level performance in grammatical error correction with hybrid machine translation." arXiv preprint arXiv:1804.05945 (2018).
[45] Kantor, Yoav, et al. "Learning to combine grammatical error corrections." arXiv preprint arXiv:1906.03897 (2019).Kantor, Yoav, et al. "Learning to combine grammatical error corrections." arXiv preprint arXiv:1906.03897 (2019).
[46] Grundkiewicz, Roman, Marcin Junczys-Dowmunt, and Kenneth Heafield. "Neural grammatical error correction systems with unsupervised pre-training on synthetic data." Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. 2019.
[47] Choe, Yo Joong, et al. "A neural grammatical error correction system built on better pre-training and sequential transfer learning." arXiv preprint arXiv:1907.01256 (2019).
[48] Li, Ruobing, et al. "The LAIX Systems in the BEA-2019 GEC Shared Task." Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. 2019.
[49] Sylviane Granger. 1998. The computer learner corpus: A versatile new source of data for SLA research. In Sylviane Granger, editor, Learner English on Computer, pages 3–18. Addison Wesley Longman, London and New York.
[50] Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto. 2011. Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), pages 147-155.
[51] Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. 2012. Tense and Aspect Error Correction for ESL Learners Using Global Context. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 198-202.
[52] Yong, Hongwei, et al. "Gradient centralization: A new optimization technique for deep neural networks." *European Conference on Computer Vision*. Springer, Cham, 2020.
[53] Soni, Madhvi, and Jitendra Singh Thakur. "A systematic review of automated grammar checking in English language." *arXiv preprint arXiv:1804.00540* (2018).
[54] Madhvi Soni, Jitendra Singh Thakur. (2021). A Comparative Analysis of Automated Grammar Checking Techniques. *Mathematical Statistician and Engineering Applications*, *70*(1), 19–26. Retrieved from https://philstat.org/index.php/MSEA/article/view/1925
[55] Jitendra Singh Thakur, M. S. . (2021). A Systematic Review of Automated Grammar Checking in English Language. *Mathematical Statistician and Engineering Applications*, *70*(1), 860 –. Retrieved from https://philstat.org/index.php/MSEA/article/view/2927

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details