

# A Research on Grammar Correction Tool with Casual to Formal Sentence Conversion

Mrs. Pallavi N. Shejwal<sup>1</sup>, Shubham Barge<sup>2</sup>, Prajwal Deokar<sup>3</sup>, Om Bothe, Gauri Ade<sup>4</sup>

<sup>1,2,3,4</sup>*Information Technology Department, PES's Modern College of Engineering, Shivajinagar, Pune.*

**Abstract:** Most of the times, we as human tend to mistakes in different fields. Talking about the correction we need a tool which can help the people to correct it accordingly. GramFormer basically deals with correction of words by suggesting the next words which are grammatically correct and are constructed properly. Apart from that it also consists of a kind of conversion of sentence from Casual to formal. In daily life, we tend to write sentences which are correct but are in informal way. But in corporate word we cannot use that or some might feel bothered to write it correctly. Overall, GramFormer provides a system to the user to grammatically correct with proper sentence construction. The fundamental of this model is based on neural network and Representation Learning.

The words a person chooses to use are a key indicator of their sense of style. Although everyone expresses themselves in a unique way, people adapt their speech and writing to the social setting, the audience, the interlocutor, or the level of formality of the situation. Text style transfer is the process of adapting and/or changing the stylistic approach used to write a sentence while maintaining the original sentence's meaning.

**Index Terms:** Natural Language Processing, Deep Learning, Sentence Conversion, Neural Networks.

## INTRODUCTION

The key to fostering positive relationships with others and conducting organised communication is effective communication. It may result in a productive workplace that is healthy, which will eventually increase productivity. Therefore, it's important to convey our demands clearly to others in order for them to respond appropriately. People often converse casually with professionals due to the change in people's lifestyles.

People frequently struggle to compose proper phrases. There may be a number of causes behind that. Usually, a lot of people write precisely correct sentences

without any mistakes. Writing with clarity, economy, and elegance does take effort. A statement that is inaccurate can give off an unexpected impression and result in unanticipated outcomes. People may therefore require a tool to address this, and natural language processing can be of assistance.

We can utilise NLP to make this issue less complicated. It is a subset of Artificial Intelligence, also known as AI, and stands for Natural Processing Language. It offers computers the ability to comprehend text or spoken words in the same way that people can.

Around the world, sentences are written in a variety of styles depending on the area, age, and gender. We therefore require a tool that can break down the linguistic barrier between them in order to successfully communicate.

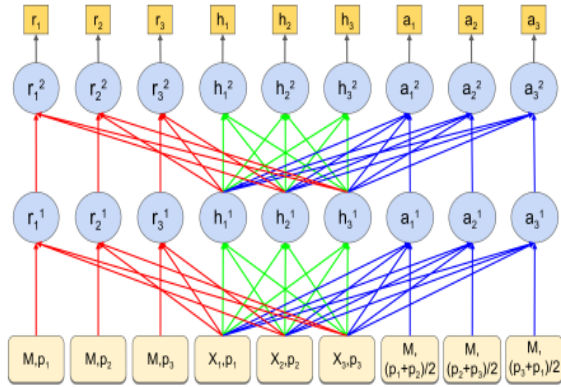
## RELATED WORK

The foundation of GramFormer is to provide a platform for the conversion of informal sentences into formal ones using Natural Language Processing and to fix English in the proper manner using neural networks. To improve the model and obtain more accurate results, experts in the field of machine learning are working on this grammatical correction. Additionally, this interface will change informal sentences into professional ones.

At first, it used statistical phrase-based machine translation (SMT) techniques that employed conventional statistical approaches.

Different models to which we have gone through are explained below.

## PARALLEL ITERATIVE EDIT (PIE)



In order to address the issue of local sequence transduction that occurs in tasks like grammatical error correction, we introduce the Parallel Iterative Edit (PIE) model (GEC). Recent methods for learning sequences from sequences are based on the well-liked encoder-decoder (ED) concept. The sequential decoding of the ED model makes it slow, but it auto-regressively captures the complete dependency among output tokens. Despite giving up the benefit of modelling full dependency in the output, the PIE model does parallel decoding and yet achieves accuracy comparable to the ED model for four reasons. In order to use pre-trained language models like BERT, one should forecast edits rather than tokens, label sequences rather than generate them, iteratively refine predictions to capture dependencies, and factorise logits over edits and their token argument. tests on a variety of tasks, including GEC, OCR correction, and spell checking

**Transformer model**

A neural network called a transformer model follows relationships in sequential input, such as the words in this sentence, to learn context and subsequently meaning.

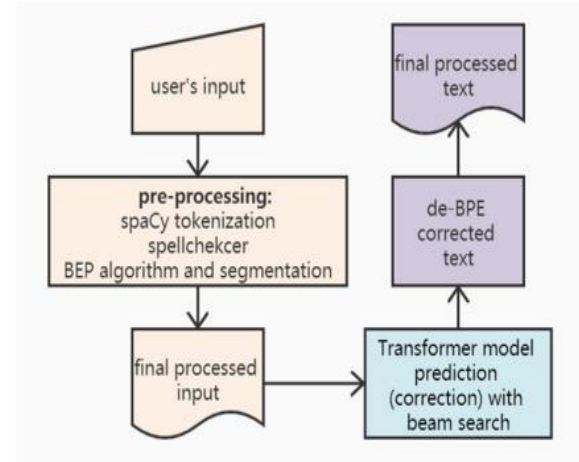
Transformer models use an expanding collection of mathematical approaches known as attention or self-attention to find minute relationships between even far-flung data pieces in a series.

Transformers are one of the newest and most potent kinds of models created to date. They were first introduced in a 2017 study from Google. They are at the forefront of a wave of machine learning innovations that some have called transformer AI.

In an August 2021 paper, researchers from Stanford referred to transformers as "foundation models"

because they believed that they would lead to a paradigm shift in AI. Our perception of what is possible has been stretched by the "sheer scale and scope of foundation models over the last few years," they wrote.

**Byte Pair Encoding Algorithm**



A straightforward method of data compression known as byte pair encoding or diagram coding substitutes a different byte for the most frequent pair of successive bytes in the data. To recreate the original data, a table of replacements is needed.

**Autoencoders**

The input and output of feedforward neural networks that use autoencoders are identical. They reduce the input's dimension before using this representation to recreate the output. The code, also known as the latent-space representation, is an efficient "summary" or "compression" of the input.

Encoder, code, and decoder are the three parts of an autoencoder. The input is compressed by the encoder, which also creates a code. The decoder then reconstructs the input exclusively using the code.

**Formality Datasets-**

1. Pavlick and Tetreault's (Pavlick and Tetreault, 2016) initial investigation on formality identification created datasets of formal and informal from news articles, emails, blogs, and local answering services. The formality of the phrases was assessed.
2. A dataset known as GYAFC has been proposed in for the research of formality style transmission for the English language (Rao and Tetreault, 2018). For training, fine-tuning, and testing, the dataset consists

of three split sets with varied data quality. The training dataset consists of 53 000 pairs of informal utterances and their translated formal equivalents. The same procedure, but with stricter quality control, was utilised to produce tuning and testing.

3. The X-FORMAL dataset, the first multilingual dataset with formality annotation, was recommended by the authors of (Briakou et al., 2021). The dataset, which also contains Brazilian, Italian, and French languages, is structurally similar to the English dataset

GYAFC. The collection has 112 thousand pairs for each language.

4. The earliest research on GYAFC (Rao and Tetreault, 2018) and X-FORMAL (Briakou et al., 2021) offered thorough experimental results using these datasets, but they were all centred on the style transfer setting and did not look into the formality detection (as a text classification task). Instead, we investigate text classification using these datasets.

LITERATURE SURVEY

Sr. No	Reference Paper (Write Paper Title)/ Field Visit (Field details)/Other Sources (appropriate source details)	Seed Idea/ Work description	Problems found / Research gap / Work gap	Your solution / Comment
1	Usage of Grammarly – Online Grammar and Spelling Checker Tool	The purpose of the current study is to ascertain the beliefs, usage patterns, and levels of satisfaction among Grammarly users. The research also discusses the methodologies utilised to create the tool.	Many people are unaware of tools like these, which may have an impact on students' growth and development. Additionally, the accuracy scores poorly compared to some other tools.	The use of Grammarly and its advantages among postgraduate students require regular orientation programmes. Additionally, group mail should be used to distribute the online information booklet.
2	A Research on Online Grammar Checker System Based on Neural Network Model (by Senyue Hao and Gang Hao)	The various models used to fix grammar are discussed throughout the entire study. We frequently have a tendency to compose grammatically wrong statements without realising it. We correct those sentences using the CNN model, the Transfer model, and alternative language. With the help of datasets.	Initial models include Parallel Iterative Edit (PIE) and statistical phrase-based machine translation (SMT) techniques using conventional statistical approaches. This has to be modified because it doesn't produce the desired outcomes.	Recent innovations like the Transfer Model and Byte Pair Encoding Algorithm can be frequently used in conjunction with various tools like SpellChecker and spaCy tokenization to obtain accurate results.
3	A Review of Text Style Transfer using Deep Learning	The review is based around two crucial steps in the text style transfer process, namely, representation learning and sentence creation in a new style. Encoder-decoder-based architectures are used.	Although it seems that choosing one deep neural network over another is style-neutral, researchers continue to face difficulties in reconciling the trade-offs between a model's complexity and the anticipated performance increases provided by auxiliary components (such as classifiers and discriminators).	A deeper knowledge of what style markers are recorded and acquired by neural networks may shed light on the nature of stylistic variances in language. Interpretability is a persistent difficulty that is shared throughout various domains.
4	Detecting Text Formality: A Study of Text Classification Approaches	The informal sentences in this paper are converted into formal sentences utilising two different datasets. Recurrent neural network and convolution neural network-based models are employed.	Cross-lingual models perform less accurately than monolingual ones.	The dataset contains samples that have been incorrectly labeled, which may be the cause of the proposed approaches' approaches' poor performance.

CONCLUSION

In recent trends, due to advanced technology it has led us to study about this particular topic with enthusiasm. During survey we got to know about different problems these models faced. By performing certain algorithms, these mistakes can be removed slightly which is definitely going to increase the efficiency of the system. Accurate use of grammar and writing is an essential aspect of any writing. Today various online grammar spelling checker tools are available for improving writing quality. Talking about the models which are used is extension of Deep Learning. It is basically a subset of Machine Learning which contains layers of neural networks. We immensely hope that this paper would help who want to study it further and want to excel in it. Transformer Model and Autoencoders are the fields that can make progress together and can provide good platform the user. Neural network is vast to learn but apart from that it helps World to improvise various things and to get accurate results in less time.

#### REFERENCES

- [1] A Research on Online Grammar Checker System Based on Neural Network. Senyue Hao and Gang Hao 2020 J. Phys.: Conf. Ser. 1651 012135
- [2] A Review of Text Style Transfer using Deep Learning. Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University - Skopje, North Macedonia
- [3] Usage of Grammarly – Online Grammar and Spelling Checker Tool Gain, Anita; Rao, Mahabaleshwara; and Bhat, Shivananda, "Usage of Grammarly – Online Grammar and Spelling Checker Tool at the Health Sciences Library, Manipal Academy of Higher Education, Manipal: A Study" (2019). Library Philosophy and Practice.
- [4] Detecting Text Formality: A Study of Text Classification Approaches Daryna Dementieva<sup>1</sup>, Ivan Trifinov<sup>2</sup>, Andrey Likhachev<sup>3</sup>, and Alexander Panchenko<sup>1</sup>
- [5] Deep Learning for Text Style Transfer: A Survey-2021 Di Jin\*, Zhijing Jin, Zhiting Hu, Olga Vechtomova, Rada Mihalcea