TrendScribe: Design and Development of a Pedagogic Trend Description Generator for Learners of English

John Blake(✉), Peng Zhao, and Evgeny Pyshkin

University of Aizu, Aizuwakamatsu 965-8580, Japan
jblake@u-aizu.ac.jp

Abstract. Times series descriptions often accompany graphs to help readers focus on the key trends. Major English language proficiency tests, such as IELTS and TOEIC, incorporate such descriptions into their written examinations. Trend descriptions are, therefore, a high-stakes genre for learners of English aiming to study at universities in Anglophone countries. To help learners become more familiar with the genre and to provide language practice at an appropriate level, we developed TrendScribe. This is the first interactive online tool that enables users to generate textual descriptions of single-line graphs from user-submitted time series data. Both rule-based and LLM-based systems are used to generate textual descriptions. Complex datasets are preprocessed using a smoothing algorithm. Users can view descriptions at their preferred proficiency level, with each level offering a corresponding increase in lexical and grammatical complexity.

Keywords: Natural language generation · Time series description · Trend description

1 Introduction

The ability to accurately describe statistical trends is a critical skill in academic and professional environments. To describe trends appropriately, it is necessary to follow the expected rhetorical conventions and harness the specific lexical and grammatical features. Describing statistical trends in English poses significant challenges for non-native speakers at lexical, grammatical and discoursal levels. Figure 1, which shows a side-by-side comparison between a model and a learner-created time-series description, illustrates the difficulty of the task.

Although numerous sources list relevant expressions [7,17], there is a paucity of accessible model texts tailored to various proficiency levels. This gap is compounded by the need for frequent, focused practice, which is essential for mastery and success in the writing sections of high-stakes language proficiency tests, such as IELTS and TOEFL. A particularly notable problem exhibited by test candidates is the limited grammatical and lexical range, particularly the lack of

C. Sombattheera et al. (Eds.): MIWAI 2024, LNAI 15431, pp. 89–101, 2025.
https://doi.org/10.1007/978-981-96-0692-4_8

The graph shows the changes in the price of a kilogram of bananas over twelve months. Overall, the price fluctuates, showing both increases and decreases before stabilizing towards the end of the year. Initially, the price starts at $1.40 and gradually increases, reaching $1.45 in February. It then stabilizes at around $1.40 for a short period before dropping slightly to $1.35 in April. After this, the price climbs steadily, peaking at $1.60 in June. Following the peak, the price decreases to $1.50 in July and then continues to drop, hitting the lowest point of $1.10 in September. After reaching this low, the price rises slightly to $1.20 and remains stable at $1.20 for the last quarter.	This graph shows one big going up and one big going down. In January to February, it gradually go up a little. Oppositely, between February and April, it gradually go down a little. However, from now on to June, it suddenly go up the most in the year. While, until September, it go down the most in the year. In September to October, it go up a little. In October to December, then it doesn't go up and go down. It won't change anything for a period of time. In conclusion, price of bananas is the highest in the year on June, also it go down for some time and it become the lowest in the year on September.

Fig. 1. Model description and a learner-created description.

sentence variety and reliance on basic trend verbs, such as *increase, rise* and *climb*. Difficulty with expression is not limited to second language learners of English, this is the case for students in the sciences and the liberal arts who need to work on their language and rhetorical skills [11].

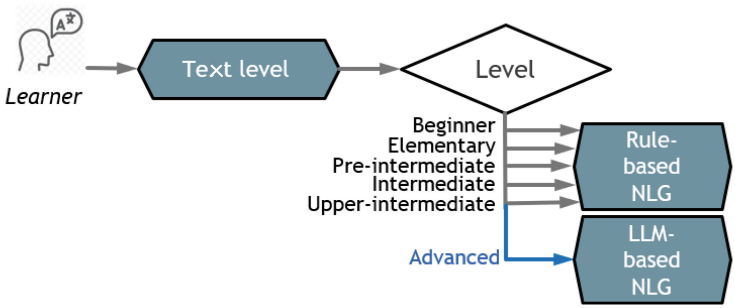

Fig. 2. Selection between the rule-based and LLM-based NLG modules

To address this problem, we propose a pedagogic trend description generator, *TrendScribe*, which creates both exemplars and practice activities. This tool is accessible online, and utilizes Natural Language Generation (NLG) to create texts and activities on demand. As Fig. 2 shows, six proficiency levels are addressed. For the first five levels, a rule-based NLG system is employed, which provides structured and predictable learning experiences ideal for foundational learning. For learners at the advanced level, an LLM-based NLG system is used, which introduces more complexity and variability in the texts generated, simulating more closely the variety found in natural language use.

Students can choose to generate full texts or engage in text-based learning activities, such as gap-fill exercises. Through these features, TrendScribe aims to provide a comprehensive and adaptable learning tool that supports learners in developing their ability to write precise and well-organized trend descriptions, thereby enhancing their linguistic competence.

This paper presents the development and initial evaluation of the TrendScribe, detailing the system and its modules, and discussing its effectiveness based on user feedback and preliminary performance metrics.

2 Related Works

2.1 Generative Language Tools

A wide range of Natural Language Generation (NLG) tools have been developed to support second language learning in English. Anthony [4] developed AntLister, a data-driven tool, which extracts high-frequency vocabulary lists and exemplar sentences from a corpus. Lee and Seneff [5] created a tool that corrects ungrammatical sentences while Rudzewitz et al. [13] created a tool that generates feedback for language exercises. Vu Tran and Blake [16] created a pedagogic tool that generates open, closed or tag questions based on an input sentence. Stowe et al. [15] also worked on question generation emphasizing control over language proficiency levels and argument structure while Morón et al. [8] used linguistic information to generate questions and answers in their version. Although not pedagogic, Obeid and Hoque [9] used a new dataset and a neural model based on an enhanced transformer-based encoder-decoder architecture to automatically generate concise, coherent natural language summaries for charts. Their model relies on the use of neural NLG to populate templated text frames.

The task of automatic generation of trend descriptions in English has been harnessed as a learning vehicle in programming classes for computer science majors [2,12] at a Japanese university. However, this research focussed on improving programming skills and developing language awareness while solving coding problems.

2.2 Research Niche

Although progress is being made in the use of neural NLG to create data series description, scant attention has been paid to the creation of texts at different proficiency levels to aid the language development of learners of English. There is also a dearth of AI-generated trend description-related activities. To the best of our knowledge, there is no other tool that language learners can use to generate pedagogic trend descriptions for language learning purposes. Our tool, TrendScribe, therefore, aims to fill this niche.

3 System Description

3.1 System Architecture

Learners input time series data values in TrendScribe manually or upload a CSV file. They then select to generate a graph, description or both. Those who want to understand the language necessary via examples, read and analyze the descriptions while learners who want to practise writing may only generate a graph, draft their version and then compare it with the generated text. Learners can view multiple textual descriptions at each of the six proficiency levels and can switch between levels to see how limiting and extended the vocabulary and grammatical range impacts the description.

Figure 3 presents a detailed view to the overall system architecture described in the following subsections.

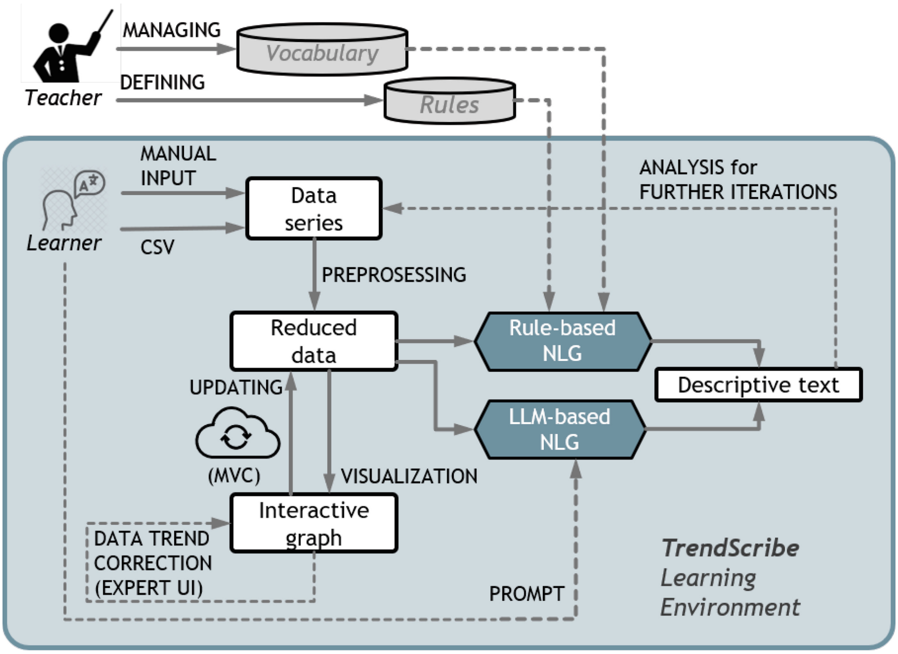

Fig. 3. Detailed TrendScribe architecture

The workflow starts with pre-processing, where complex datasets are reduced using smoothing algorithms to enable concise descriptions to be generated. Following this, the values in the dataset are analyzed to identify: general and specific trends, inflection points, maxima and minima, the degree of change and so forth, each of which contributes to the natural language generation process. The user-friendly GUI provides users with a submission form to upload or input data values, visualize the resultant graph, and display the generated text. To synchronize the data updates with the possibilities to edit the graph, the standard model-view-controller architectural pattern (MVC) [6] can be used.

3.2 Pre-processing Module

The pre-processing module handles complex raw data inputs, typically in formats such as CSV, and applies a smoothing algorithm to reduce the number of data points to a manageable quantity. This simplification is necessary for generating understandable and relevant textual descriptions without overwhelming users or the NLG system.

The data graph is smoothed using the Savitzky-Golay method [14] provided in the Python library SciPy. This technique employs weighted moving averages between the peaks and valleys of the graph through setting the degree of the polynomial and the window size, which need to be specified beforehand. The window size indicates the number of variables considered up to n points in the past, including the current point, thus it should be $2n + 1$ for n points. For example, a polynomial of degree 3 is shown in Eq. 1.

$$y = a_3x^3 + a_2x^2 + a_1x + a_0 \tag{1}$$

Increasing the window size results in a smoother graph, while a higher polynomial degree highlights changes. By combining these parameters, the graphs are approximated to mimic human expertise. The determined numbers are used for the least squares method within the Savitzky-Golay method.

$$u(x) = \sum_{n=0}^{N} x^n a_n \tag{2}$$

In Eq. 2, $u(x)$ is the smoothed value, a_n is the coefficient of the n-th order term, and x is the window size [18]. By differentiating this value, the noise strength in the line graph can be determined. Figure 4 shows two examples illustrating the possibility to smooth the data in order to emphasize the larger regions focusing the meaningful trends depending on the desired sensitivity to the data deviations. To produce these examples, we used the Japanese gasoline sales prices between Aug 2019 to Jun 2023 available in [3].

3.3 Rule-Based NLG Module

We assume that the proficiency level in the target language influences a language learner or user's ability in vocabulary selection, sentence construction, and overall comprehension of the main ideas in writing. The ability discussed here does not merely refer to the distinction between correct and incorrect usage, but rather encompasses the breadth of vocabulary choice, the richness of sentence

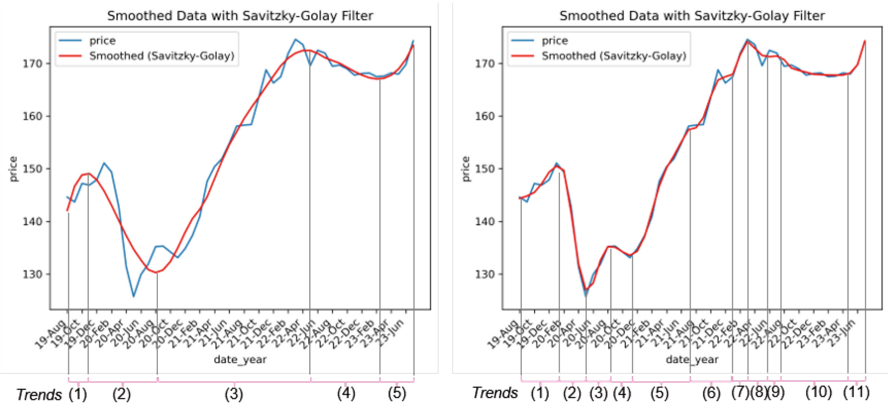

Fig. 4. Examples of data graph smoothing with different sensitivity to deviations

structure, and the clarity of the overall message. Based on this understanding, our rule-based NLG, while ensuring grammatical correctness, attempts to generate exemplar descriptions tailored to learners of different proficiency levels by adjusting the linguistic complexity at lexical, grammatical and discoursal levels.

For each language proficiency level, pre-established multiple candidates were selected for rule-based NLG. The descriptive text presented to users, as a result, is generated based on a series of predefined rules, rather than a single pattern, corresponding to the selected language proficiency level. Consequently, each generation result employs varied sentence structures, vocabulary, and other linguistic elements, aiming to demonstrate the richness of the generated content and convey this linguistic flexibility to the user-learners. Table 1 demonstrates a selection of examples.

Table 1. Adjusting the text description to fit the language proficiency levels: Examples

Levels	Lexical	Grammatical	Discoursal
Beginner	verbs	simple sentences	specific trends
Elementary	nouns	combine sentences	general trend
Pre-intermediate	prepositions	extend sentences	noun vs verb style
Intermediate	adverbs	complex sentences	noun vs verb style
Upper-intermediate	all	sentence variety	cohesive devices

Beginner Level. For beginners, writing grammatically correct sentences is crucial. The key requirement is to describe each data change using one correct sentence, which combine to form a paragraph. Ideally, beginners are also expected to avoid excessive repetition of words, instead alternating between words or syntactic structures with similar meanings, i.e., employing synonymous substitutions. The system takes this into account when generating the sample texts for beginner level. For instance, although *increase* is an appropriate word for describing data growth, the system uses *climb, rise*, or *growth* to replace some instances of *increase* to enhance expressive variety. Similarly, in addition to noun-style trend phrases like *There was a rise*, the system also generates verb phrases such as *The value rose* to introduce the changes of data.

Elementary Level. Learners at the elementary Level are expected to convey effective information more concisely and appropriately. Specifically, in trend description tasks, compared to the systematic sentence stacking for beginner level learners, elementary Level learners are considered capable of observing the overall direction of data changes and generating appropriate descriptions for each change direction. For example, when data consistently increases from January to March, elementary Level learners are expected to describe this trend in one sentence rather than multiple sentences. This simplification of complex information facilitates more efficient delivery of key information, demonstrating an improvement in ability.

Pre-intermediate Level. For pre-intermediate level learners, it is important to grasp the data changes over the entire time period more comprehensively. Learners at this stage are expected to further free their attention from simple, repetitive sentence assembly operations, focusing on data change trends in various directions while better grasping the overall picture of the data. The sample texts generated by the system for pre-intermediate level learners also take this into account, adopting various methods to demonstrate ways of describing the overall data picture. Specifically, when the annual data shows a growth trend, the system generates diverse summary trend descriptions, such as the verb-style sentence *Over the year, the value dropped from 110 to 80 by 30.* or the noun-style sentence *Over the year, there was a decrease from 110 to 80 by 30.*

Intermediate Level. Grammar and meaning are two indispensable branches of language learning. At the intermediate level, a key challenge is sentence construction. In other words, intermediate level learners, compared to those in the previous two stages, face greater challenges in enhancing the richness of their sentence structures. They will attempt to use more adjectives or adverbs to modify and describe the degree of data changes, and longer sentences once again bring the issue of sentence correctness to the forefront. The system adapts to this pattern by incorporating adjectives or adverbs into sentences generated for the intermediate level, providing students with examples of more richly constructed sentences. Instead of *There was a rise from 110 by 50 from January to May*, the generated text will create sentences like *The value rose negligibly from 110 by 50 from January to May.* to create more descriptive texts.

Upper Intermediate Level. For upper intermediate level learners, after mastering the methods of using adjectives and adverbs to modify sentence components, they can correctly and proficiently enrich their sentences. In terms of enriching meaning, such learners are expected to provide concise, clear, and varied descriptions of the maximum and minimum values across the entire range of change. Therefore, the system incorporates diverse descriptions of the overall range's extremes, generating sentences like: *The value bottomed out at 10 in August, and peaked at 160 in May.*, and *The value hit a low of 10 in August.*

3.4 LLM-based NLG Module

The system is a ring-fenced Large Language Model designed to generate textual descriptions of time series data using chain-of-thought prompt engineering, leveraging multiple advanced technologies and libraries for optimal performance. At its core is the Phi-3 Medium, a 14 billion parameter transformer-based language model developed by Microsoft, which generates detailed descriptions for time series data. To enhance the quality of these examples, Gemini Advanced is employed for quality control. Ollama serves as the inference engine, enabling the efficient execution and management of AI models. Additionally, Langchain facilitates the integration of these large language models for various natural language processing tasks, managing data flow and connecting different components to ensure seamless operation. This agentic system combines these technologies to produce high-quality, efficient, and accurate outputs, serving as exemplar texts for advanced learners.

3.5 Graphical User Interface and User Experience

The system design adheres to the design principles of simplicity and directness, aiming to enable users to quickly understand what to do when interacting with the interface. As can be seen in Fig. 5, the viewport is divided into three areas: a title field, a generated results output area, and a data input area.

In the generated results output area, the system explicitly reserves space for both the text-based output and the chart-based output, which serve as visual supplements to the textual explanations, reducing the complexity of comprehending lengthy text passages. Reserved spaces in this area are designed to provide users with a preliminary understanding and expectation of system output. In the data input area, users can either input values via a submission form or upload correctly-formatted CSV files to achieve one-time batch data input.

The title field occupies only a small space at the top of the page, allowing users to input the subject of the submitted dataset. The user-defined subject will be utilized in the generation process of the data analysis text and reflected in the final output. The generated results output area comprises two sections: a textual description and graphical visualisation.

In order to start the generation of sample text, users can click on one of six predefined green buttons labeled: *Beginner, Elementary, Pre-intermediate, Intermediate, Upper Intermediate,* and *Advanced.* The system then generates

Fig. 5. GUI showing output

and displays a text description in the designated area based on the user input data, adhering to predefined rules corresponding to the selected language proficiency level. The chart output section, which is located below the text output section, supports the display of system-generated line graphs that can show data trend changes effectively. For users, the graphical representations provides them with a valuable supplement to the textual analysis, enhancing their overall comprehension towards data and the description of its trend. Moreover, even when users have not initiated any text or chart generation, the generated results output area still occupies a space at the top of the page, displaying simple and easy-to-understand operational guidelines within this area. This design stems from our consideration of the users' needs. We believe that for users, constantly being aware of the system's generation output functionality and having convenient access to the latest generated results is beneficial for improving their operational efficiency and learning effectiveness.

The user's interaction begins with the data input area. Upon page loading, the system populates this area with default values. This design not only serves as a demonstration for users but also allows them to experiment with the system's generation functionality without manually inputting values, should they wish to explore and understand the system's capabilities as quickly as possible. User data input on the page is subject to limitations. Specifically, users are not allowed to input more than 12 time points and their corresponding data items. For simple "trend description writing" exercises, an excessive amount of data would introduce unnecessary complexity, potentially undermining the primary purpose of describing trends. However, when necessary, users can still use the *Upload CSV file* feature to batch upload large amounts of data at once. This means that while direct, page-based data input is limited, the system still supports operations involving large-scale data input.

4 Preliminary Evaluation

4.1 Usability

The usability of TrendScribe was evaluated through feedback received from computer science majors enrolled in a credit-bearing elective course on Technology-Enhanced Language Learning in a Japanese university. The feedback focused on user satisfaction and interface effectiveness, with students providing specific advice on how to enhance the overall user experience. Actionable feedback received included (1) suggestions regarding the generated text, such as capitalization issues; (2) comments regarding the positioning of the elements, specifically placing the generated text above the data input areas for ease of use, and (3) generating a graph, since the initial version only displayed the data values.

4.2 Accuracy

In assessing the accuracy of the generated text by the TrendScribe, we adopted a benchmark for natural language generation, focusing on markedness, which involves the frequency of use and adherence to accepted linguistic conventions. Markedness [1] is used to differentiate frequent preferred unmarked forms from the less frequent marked language forms, which may or may not be grammatically correct.

Two human raters evaluated the output of TrendScribe at textual and sentential levels. First, a holistic evaluation of each generated complete text was conducted, focusing on organization, coherence, and cohesion. All texts generated were suitable in terms of organization, coherence and cohesion. Second, a discrete evaluation of the markedness of each sentence was undertaken, classifying it as either marked or unmarked. All the sentences generated by the LLM were unmarked. However, approximately 15% of the rule-based sentences contained marked language. This is, however, easy to rectify in the next release.

4.3 Efficacy

Pilot studies of TrendScribe with a focus group of three university-age Japanese learners of English resulted in positive feedback. Users were observed interacting with the GUI in the expected manners, viz: using the tool to generate examples to read or to generate graphs to describe and then compare their draft to the generated model. The users noted that they learnt some new sentence structures and and new vocabulary. More systematic testing is needed, however, to establish whether TrendScribe can help learners understand and create trend descriptions more effectively compared with other methods of study.

5 Conclusion

TrendScribe can generate exemplar texts at six different language proficiency levels from user-submitted data. These generated texts serve as a practice vehicle for students who need to describe graphs. The texts provide contextualized examples of the prototypical use of both lexis and grammar, helping learners master this challenging sub-genre of describing trends.

The two main limitations of the current release are that (1) some rule-based output is either ungrammatical or marked, and (2) learners may not be aware of the type of language that is being focussed on at each level. To address these limitations, further refinements need to be made to the rule-based NLG algorithms to ensure that the output is always grammatically correct and pitched at the appropriate level. To make the workings of the algorithm more transparent, and to help learners understand what language features are being focussed on, details of the language under focus will be shared with users.

There are many potential expansions for TrendScribe, including the addition of other practice activities to help learners master the genre of trend descriptions. As we suggested in [10], an expert UI can be helpful for marking up the trend-representing regions directly on the displayed graph, in the case when the automated data smoothing algorithm applied to the real-world data does not produce completely satisfactory regions from the perspective of their suitability for language learning purposes.

Additionally, future work will include a more rigorous evaluation using pre- and post-test comparisons to measure the impact of TrendScribe on learning outcomes quantitatively, and ideally adopting an experimental approach.

Acknowledgments. This study was funded by JSPS Grant-in-Aid for Scientific Research (Grant number 22K00792). We would like to acknowledge the work on the smoothing algorithm by Mr. Yuki Matsuzaki under the supervision of Prof. Pyshkin.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Blake, J.: Incorporating information structure in the EAP curriculum. In: Conference Proceedings of 2nd International Symposium on Innovative Teaching and Research in ESP (2015)
2. Blake, J., Tamura, K., Pyshkin, E.: Linguistic-first approach to learning python for natural language generation: Problem breakdown to pseudocode. In: AIP Conference Proceedings, vol. 2909. AIP Publishing (2023)
3. General Incorporated Association Energy Information Center: National gasoline prices (yen/liter). https://pps-net.org/oilstand Accessed Sep 2024
4. Laurence, A.: Automatic generation of high-frequency vocabulary lists and level-adjusted exemplar sentences for non-native speakers of English. In: Proceedings of the 41st Annual Meeting of the English Language Education Society of Japan (2011)
5. Lee, J., Seneff, S.: Automatic grammar correction for second-language learners. In: Interspeech 2006. ISCA (Sep 17 2006)
6. Leff, A., Rayfield, J.T.: Web-application development using the model/view/controller design pattern. In: Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference, pp. 118–127. IEEE (2001)
7. Morley, J.: Academic phrasebank. University of Manchester, Manchester (2014)
8. Moron, M., Scocozza, J., Chiruzzo, L., Rosa, A.: A tool for automatic question generation for teaching English to beginner students. In: 2021 40th International Conference of the Chilean Computer Science Society (SCCC). IEEE (Nov 15 2021)
9. Obeid, J., Hoque, E.: Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model. arXiv preprint arXiv:2010.09142 (2020)
10. Panda, S., Gomez, F.P., Flor, M., Rozovskaya, A.: Automatic generation of distractors for fill-in-the-blank exercises with round-trip neural machine translation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 391–401 (2022)
11. Pyshkin, E.: Liberal arts in a digitally transformed world: Revisiting a case of software development education. In: Proceedings of the 13th Central & Eastern European Software Engineering Conference in Russia, pp. 1–7 (2017)
12. Pyshkin, E., Blake, J.: Bringing linguistics to a programming class: A problem of automatic text generation for describing data series. In: New Trends in Intelligent Software Methodologies, Tools and Techniques, pp. 621–630. IOS Press (2022). https://doi.org/10.3233/FAIA220291
13. Rudzewitz, B., Ziai, R., De Kuthy, K., Möller, V., Nuxoll, F., Meurers, D.: Generating feedback for English foreign language exercises. In: Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications. Association for Computational Linguistics (2018)
14. Savitzky, A., Golay, M.J.: Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. **36**(8), 1627–1639 (1964)
15. Stowe, K., Ghosh, D., Zhao, M.: Controlled language generation for language learning items. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track. p. 294-305. Association for Computational Linguistics (2022)
16. Vu Tran, D., Blake, J.: Design and development of a question generator for learners of english. In: SHS Web of Conferences. vol. 102, p. 01011. EDP Sciences (2021)

17. Wilson, J.: Recent IELTS materials. ELT J. **64**(2), 219–232 (2010)
18. Yoshimura, N., Takayanagi, M.: Chemometric calculations using Microsoft Excel. J. Comput. Chem. **11**(3), 149–158 (2012)