OPA: An AI System for Automating the Correction and Analysis of Portuguese Dictations

Erika C. Matesz Bueno[1], Polyana Graf Finamor[1], Thaynara Cardoso[1], Ana Cristina Bicharra Garcia[1]

[1]Programa de Pós-Graduação em Informática –
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Caixa Postal 56.011 – 22.290.970 – Rio de Janeiro – RJ – Brazil

{erikamatesz, polyanagraf, thaynara.cardoso}@edu.unirio.br,

cristina.bicharra@uniriotec.br

Abstract. *This paper presents an OPA (Observation, Processing, and Assistance) proof of concept, an AI-based system designed to support Portuguese language teachers in public schools by automating the correction and analysis of handwritten dictations. OPA uses OCR technologies and natural language processing to digitize students' written work, identify errors, and classify them by type. This allows teachers to provide targeted interventions and better manage classroom data. By automating routine tasks, OPA aims to save time for educators and improve student literacy. The study demonstrates the potential of AI to transform educational practices, particularly in resource-limited environments, and addresses the lack of AI tools in Portuguese language education.*

1. Introduction

Despite the Digital Age we're in, the Dictation method still remains an effective way to diagnose grammatical errors in writing [Xuechen 2009] [Luo et al. 2022]. It evaluates both students' listening skills and written production; moreover, it ensures the students will have attentive listening and helps fix punctuation concepts, which are points strongly affected by the excessive use of technology nowadays. However, the traditional ways of correcting dictation usually consist of self-correction, which places the responsibility on the student and provides little or no insight to the teacher about the learner's progress. Or, the manual correction by teachers, who typically have hundreds of students (particularly in public schools), a time-consuming and labor-intensive process that requires a systematic analysis of the learners' errors [Xuechen 2009] [Luo et al. 2022].

Hence, considering teaching is both an art and a science, this is where Artificial Intelligence (AI) steps in. Advances in AI technologies, particularly in handwritten text recognition and natural language processing, offer powerful tools to augment the work of educators. The integration of AI into educational contexts can automate tasks that would otherwise require significant manual effort, freeing teachers to focus on personalized instruction and innovative pedagogy.

In this paper, we introduce a proof of concept (POC) of OPA (Observation, Processing, and Assistance), an AI-powered system designed to support educators in public schools by addressing one of their most pressing challenges: analyzing and responding to students' writing errors. Initially conceived as a POC, the OPA system is now evolving toward a minimum viable product (MVP). It leverages state-of-the-art AI capabilities in handwriting recognition to digitize and analyze handwritten essays in brazilian portuguese. It identifies errors and classifies them by type (e.g., spelling, grammar, or syntax) up to this point; however, in a future step, it will also cluster similar errors to reveal patterns at both the class and individual levels. As part of this development, the OPA system is being designed to include the generation of a detailed textual report for teachers, highlighting the most frequent mistakes and providing tailored suggestions for alternative learning strategies. These features are currently under active development and are grounded in the analysis of identified error patterns, aiming to support teachers in designing more effective interventions, as suggested by [Negro et al. 2024].

This work demonstrates how AI can be harnessed not only to automate routine tasks but also to provide meaningful insights that empower teachers and enhance the learning process. By bridging the gap between data analysis and actionable strategies, OPA exemplifies the potential of AI to amplify the capabilities of educators and improve educational outcomes in resource-constrained settings.

2. Theoretical Reference

Generalized teaching strategies are increasingly ineffective for language learning. Studies like [Negro et al. 2024] highlight the value of individualized approaches, emphasizing error analysis, distributed strategies, and continuous feedback to enhance learning, retention, and spelling transfer. However, while advances like Intelligent Tutoring Systems (ITS) automate educational exercises by simulating human tutors, systems such as [Xuechen 2009] and [Zhai et al. 2022] web-based dictation correctors remain strictly digital, receiving inputs via keyboard and lacking interaction with human teachers and collective learning.

Our study focuses on enhancing the correction process for dictations conducted live by teachers in classrooms, presenting two essential problems: how to read the handwritten exercises submitted by students and how to efficiently correct them. These challenges have been addressed in various ways, often with limited success, largely due to the constraints of existing Optical Character Recognition (OCR) technologies and the need of large amounts of data.

A significant tool in this field is ERRANT [Bryant et al. 2017], developed to analyze grammatical errors and provide feedback to English learners. Although it marked a major step forward, its design is specific to English and was not originally intended for other languages or educational contexts. However, efforts to adapt this tool have resulted in systems like FRETA-D [Luo et al. 2022], which modified ERRANT to incorporate grammar-specific rules for French, catering to learners of French as a foreign language.

Even so, handwritten input still poses unique difficulties for OCR systems, as demonstrated by [Santos et al. 2023] and [Deepthi and Seenu 2022], which highlight the variability in handwriting styles and the complexity of character structures. For instance, while digital input methods like keyboards or Apple Pencil facilitate error identification, they overlook issues such as spacing errors around punctuation, which are more apparent in digital formats than in handwritten exercises. Moreover, studies evaluating OCR mechanisms for Brazilian Portuguese (Pt-Br) remain scarce, contrasting with more extensive research in English and Arabic [Santos et al. 2023]. Similarly, [Wang 2024] addressed challenges in Chinese character recognition by using artificial intelligence to create a comprehensive database incorporating stroke order rules and segmentation algorithms, significantly improving recognition accuracy for complex writing systems. And [Guan et al. 2020] proposed GAN augmentation data to better train their model.

Following a different path, [Wojcicki and Zientarski 2024] developed an n-gram-based model that corrects words using Damerau-Levenshtein Distance and Sorensen-Dice algorithms during the post-processing of Polish documents. This approach demonstrates significant adaptability for other languages and succeeds in processing handwritten documents. However, its reliance on a dictionary of approximately 5 million Polish words underscores the need for extensive linguistic resources to achieve accurate corrections. Similarly, [Wen et al. 2024] acknowledged that their Multi-Feature Data Fusion Algorithm works effectively but suffers from limitations due to insufficient data, leaving gaps that other systems with larger datasets could address.

Despite these advancements, implementing these methodologies requires expensive and complex computational resources. Hence, the OPA project searches for other ways to deal with these questions. Furthermore, in the mentioned studies above, only [Zhai et al. 2022] was designed as a tool for teachers (although it can also be used by learners), while all other education-focused projects exclusively targeted language learners, which sets our work apart.

3. Objective

Considering the environment of public schools and the few computational resources brazilian teachers usually have access, we propose a system that consist in using AI for our two most important factors: the OCR and the automate correct system.

At the moment, correcting handwritten dictations involves doing everything manually. In addition to preparing and delivering lessons, teachers handle multiple classes and hundreds of students each school year. Furthermore, each student completes several dictations throughout the year, and the correction of each one is individual and 100%manual. The data from these corrections must also be manually entered into spreadsheets, which, combined with the rest of the process, makes the entire workflow extremely labor intensive for these professionals.

The OPA model is proposed to automate nearly all parts of this process by implementing a POC for a system to correct dictations and extract metrics using AI. While the correction functionality is operational, the modules responsible for obtaining personal and class-level statistics are still under development. Once completed, these features will allow teachers to better manage their classes and students, enabling the reuse of student data as they progress through school years. Additionally, the model is intended to assist in creating new, fully customized dictations for individual students or entire classes.

4. Research Stage

We employed prompt engineering techniques and structured our model into three stages: OCR, automated correction and classification of students' writing errors, and the generation of statistics based on the correction results. While the first two steps have been completed, it is important to note that the statistical generation component is still under development.

4.1. Technologies

An investigation was conducted into the OCR capabilities of Amazon Textract, Google Gemini, and PyTesseract. As evidenced by the findings in [Santos et al. 2023] and illustrated in Fig. 1, PyTesseract demonstrated significantly suboptimal performance. Based on these results, despite being a free option, it was ultimately disregarded.

On the other hand, the first two tools mentioned were paid but performed as expected. And among these paid options, Amazon Textract was the most expensive. Given budget constraints, we opted for Google Gemini models to perform OCR, which offered a more cost-effective solution.

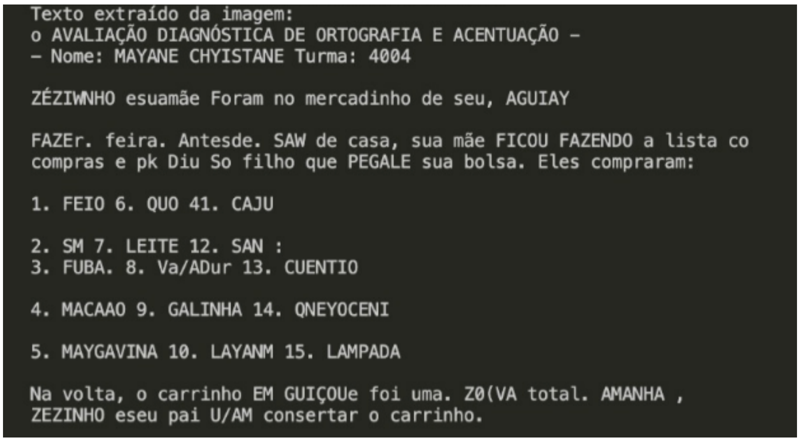

Figure 1. PyTesseract

It was also considered that OCR is just one step in the process, followed by analysis via AI. Therefore, the use of Google Gemini was considered for both the OCR stage and dictation correction, as it offers a single, simpler implementation with only one cost source.

Based on evaluation experiments, Google Gemini 1.5 Flash was selected for OCR. However, for automated correction, the Pro model was chosen, as initial tests showed that the Flash version, while effective for OCR, did not meet expectations in the correction process.

4.2. OCR

The OCR is responsible for reading students' handwritten exercises and transforming the digitized images into corresponding texts. However, different people have different handwriting styles, and, in addition, there is the possibility that the manuscripts may contain erasures. This combination of factors highlighted the need to handle a wide variety of fonts, handwriting styles, and interferences caused by corrections or manual alterations. Moreover, even humans sometimes struggle to distinguish certain letters without knowing the person who wrote them and the context in which they were written.

For these specific cases, once the OCR processes the handwritten dictation, the raw text is made available to the teacher for review and possible adjustments. This is done by displaying the digitized manuscript (a .JPEG image) alongside the OCR-generated text in an editable text box. The teacher can then compare the text with the original manuscript and make minor adjustments to ensure the OCR result exactly matches what the student wrote, without correcting actual errors in the text.

In terms of input methods, the handwritten dictations can be digitized and inserted into the system using either a cell phone camera or a scanner. There was no statistically significant difference in the OCR accuracy between these two methods, indicating that both options are viable for capturing the student's work. For a low-cost solution, using a cell phone camera offers the best cost/benefit ratio without compromising the quality of the digitization process.

4.2.1. OCR Evaluation

The AI was explicitly directed to ignore elements such as titles, names, and class identifiers, as well as to avoid making any corrections or adjustments to the transcribed text. However, it was observed that, in some cases, the AI attempted to automatically correct, complete, or infer meaning from the text, contradicting the primary objective of performing only transcription. To address this, the prompt was rewritten and further simplified to ensure the process focused exclusively on the OCR stage.

The evaluation used 12 samples, including partially printed and handwritten texts. Five were provided by the teacher involved in the POC development, while the other seven came from a similar, slightly longer dictation set.

It is important to highlight that all samples contain spelling and accentuation errors, which were intentionally preserved throughout the process, as the primary objective was to evaluate the model's transcription capability rather than its proficiency in correcting the texts.

As shown in Table 1, the results indicate an average of 12 adjustments per sample, representing 2.3% of the total text obtained through OCR.

Sample	Characters Read	Adjustments Made	% Changes
001	580	14	2.4%
002	584	14	2.4%
003	580	18	3.1%
004	584	6	1.0%
005	581	8	1.4%
006	585	14	2.4%
007	575	14	2.4%
008	426	8	1.9%
009	425	10	2.4%
010	428	8	1.9%
011	380	8	2.1%
012	383	16	4.2%

Table 1. OCR Results for Mixed Handwritten and Printed Samples

To further assess the model's consistency, an additional test was conducted using 19 fully handwritten samples. As with the previous test, these samples also contained spelling and accentuation errors. Among the analyzed samples, only one was error-free, and it was included solely to evaluate the model's transcription performance.

As illustrated in Table 2, these results show an average of 9 adjustments per sample, representing 1.8% of the text obtained through OCR.

From this OCR POC, the implementation evolved into an application, built using the Python Django framework for the back-end, along with a front-end built using HTML, CSS, and JavaScript. The system was divided into two main parts: the dictation correction mechanism and the statistics generation mechanism, as shown in Figure 2.

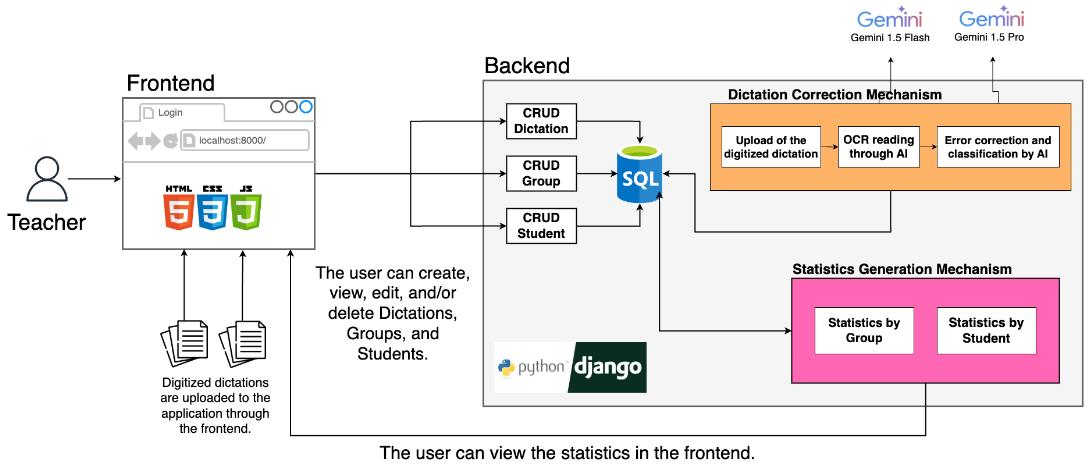

Figure 2. Prototype

4.3. Automated Correction

For AI-based correction, a table of Brazilian Portuguese grammatical rules was created, including examples of errors categorized by type. The types of errors could be categorized into macro groups of regularities and irregularities, and further classified into more specific micro groups, such as Direct Regularities, Contextual Regularities, and Morphological-Grammatical Regularities, each with their respective examples of orthographic deviations.

This table was manually developed by the team with the support of external materials provided by the Portuguese language teacher involved in the project. Using this data, the AI was trained to identify errors in the exercises, differentiate between correct and incorrect elements in the text, and classify the mistakes. This process laid the groundwork for the third stage: generating statistics.

4.3.1. Evaluation of Text Correction

Based on our table of Portuguese grammatical rules, as referenced in subsection 4.3, the Gemini 1.5 Pro model performed incredibly well, achieving an average of 90% accuracy in correcting the obtained samples, providing an excellent starting point for the evaluation tests and demonstrating the model's potential, as shown in Table 3.

The entire process was carried out through the application's interface (currently under development) as show in 3, which allows the selection of the class and the student, the selection of the dictation key (1), and the upload of the handwritten dictation with a preview (2). When starting the OCR process, the reading will be displayed in the text box (3) and can be adjusted to fully match the handwritten text. Moving on to the actual correction, the result is shown in the lower text box (4), displaying the word from the answer key (correct form), how it was written by the student, and the mistake made. If no mistake was made, it is indicated with a 0.

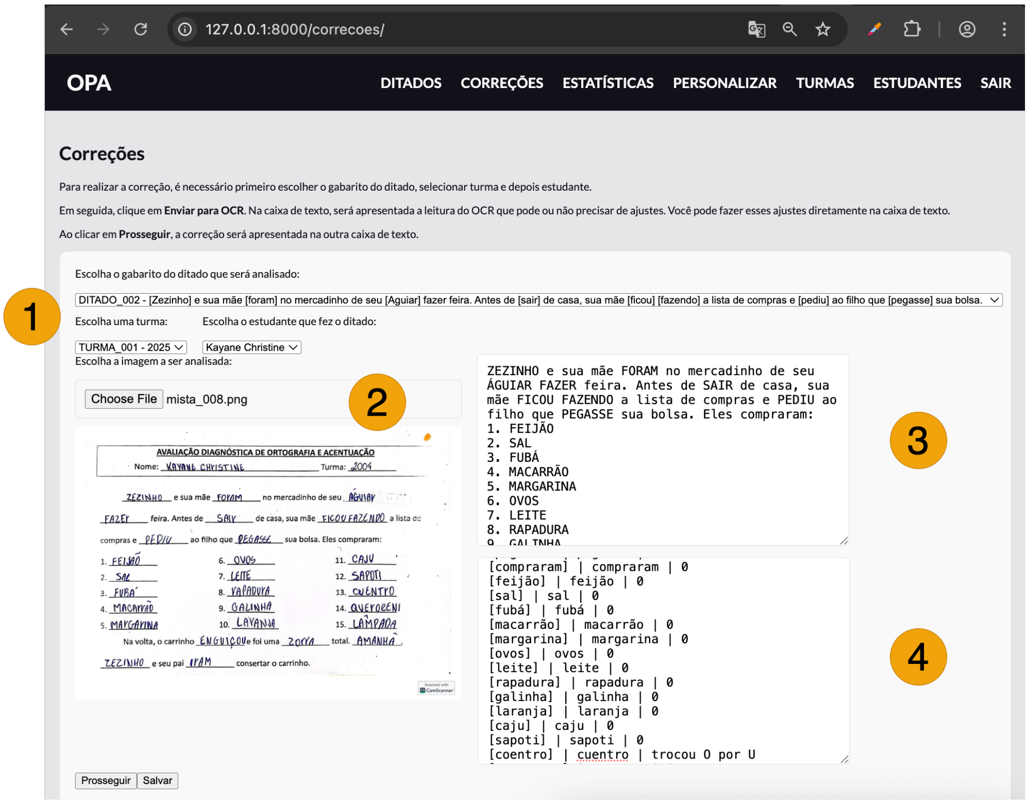

Figure 3. Correction screen inside the application

4.4. Contribution Potential

Once the statistics component is fully implemented, allowing for the storage of student and class data, Portuguese language teachers from different classes and years will be able to use these insights to improve their teaching practices. While in Brazil, information exchange between teachers usually happens informally — mainly during class councils — access to this data will allow educators to track students' progress over the years. Additionally, this exchange of information will help identify individual and collective challenges faced by students, something that is often difficult to achieve in a structured way, as many teachers do not have the time or incentive to review reports from other classes. With this data readily available, teachers will be able to plan lessons more effectively and adjust their teaching strategies to better meet students' needs, fostering greater collaboration among educators and enhancing the learning experience.

Sample	Characters Read	Adjustments Made	% Changes
001	501	4	0.8%
002	495	8	1.6%
003	499	0	0.0%
004	500	6	1.2%
005	505	16	3.2%
006	494	4	0.8%
007	496	10	2.0%
008	512	12	2.3%
009	508	6	1.2%
010	507	6	1.2%
011	502	8	1.6%
012	506	6	1.2%
013	500	10	2.0%
014	507	10	2.0%
015	497	14	2.8%
016	492	10	2.0%
017	477	16	3.4%
018	507	16	3.2%
019	493	6	1.2%

Table 2. OCR Results for Fully Handwritten Samples

Samples	Words To Be Corrected	Wrong Corrections	Accuracy
001	22	0	100%
002	22	0	100%
003	22	0	100%
004	3	0	100%
005	22	0	100%
006	21	3	86%
007	22	0	100%
008	4	1	75%
009	11	0	100%
010	7	4	43%
011	8	8	0%
012	8	2	75%
Average	14	2	90%

Table 3. Automated Correction Results

References

Bryant, C., Felice, M., and Briscoe, E. (2017). Automatic annotation and evaluation of error types for grammatical error correction.

Deepthi, C. V. S. and Seenu, A. (2022). A systematic review on ocrs for indic documents manuscripts. In *2022 International Conference on Data Science, Agents Artificial Intelligence (ICDSAAI)*, volume 01, pages 1–4.

Guan, M., Ding, H., Chen, K., and Huo, Q. (2020). Improving handwritten ocr with augmented text line images synthesized from online handwriting samples by style-conditioned gan. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 151–156.

Luo, Y., Zhai, Y., and Qin, Y. (2022). Freta-d: A toolkit of automatic annotation of grammatical and phonetic error types in french dictations. In *2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS)*, pages 531–537.

Negro, I., Leblanc, N., and Bonnotte, I. (2024). How to individualize lexical spelling instruction with distributed retrieval and feedback: an exploratory study with first-grade french students. *Reading and Writing*.

Santos, Y., Silva, M., and Reis, J. C. S. (2023). Evaluation of optical character recognition (ocr) systems dealing with misinformation in portuguese. In *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 223–228.

Wang, R. (2024). A method for constructing a dictation system based on artificial intelligence technology and a dictation machine. In *2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 302–305.

Wen, J., Feng, X., and Fu, F. (2024). English text spelling error detection and correction based on multi-feature data fusion algorithm. In *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, pages 1–5.

Wojcicki, P. and Zientarski, T. (2024). Polish word recognition based on n-gram methods. *IEEE Access*, 12:49817–49825.

Xuechen, H. (2009). A web-based intelligent tutoring system for english dictation. In *2009 International Conference on Artificial Intelligence and Computational Intelligence*, volume 4, pages 583–586.

Zhai, Y., Tian, N., and Huang, X. (2022). *Exploring the Design and Application of an Intelligent French Dictation Platform*.