BILINGUAL PUBLISHING GROUP

**Forum for Linguistic Studies**

https://journals.bilpubgroup.com/index.php/fls

**ARTICLE**

# Design of Intelligent Educational Mobile Apps with an Original Dataset for Chinese-Portuguese Translators

***Lap Man Hoi*** [1*] **, *Yuqi Sun*** [1] **, *Manlin Lin*** [2] **, *Sio Kei Im*** [2]

[1] *Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, China*

[2] *Engineering Research Centre of Applied Technology on Machine Translation and Artificial Intelligence, Ministry of Education, Macao Polytechnic University, Macao SAR, China*

## ABSTRACT

Translation remains a vital process in many culturally diverse countries. Despite significant advances in *artificial intelligence* (AI) technology, machine translation currently lacks the ability to fully replace human expertise, requiring continued human intervention and review in translation workflows. This article introduces an innovative mobile education application (app) designed to train translators, with a particular focus on Chinese-Portuguese translation. This app uses a set of practice data, *Chinese-Portuguese translation exercise corpus* (CPTEC), developed by our corpus team to autonomously assess and identify translation quality defects, thereby promoting skill improvement. We also propose a novel hybrid grade system based on different *translation quality assessment* (TQA) dimensions to automatically evaluate translations by imitating humans. In addition, it demonstrates the design of challenging exercises within a mobile app to reinforce translation proficiency. To optimize the functionality of the mobile app, we use a *large language model* (LLM) to validate the solution, ensure that it learns the training material provided and track its performance. Subsequent experimental results show that the fine-tuned LLM improves on multiple dimensions (including accuracy, fidelity, fluency, readability, acceptability, and usability) compared to the initial state, confirming the effectiveness of the developed practice data in improving translation performance. To promote access to research, the practice data (CPTEC) will be distributed within the relevant AI community, to inspire people to create innovative software applications to support translators.

*CORRESPONDING AUTHOR:

Lap Man Hoi, Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, China; Email: lmhoi@mpu.edu.mo

ARTICLE INFO

Received: 21 April 2025 | Revised: 26 May 2025 | Accepted: 9 June 2025 | Published Online: 16 July 2025
DOI: https://doi.org/10.30564/fls.v7i7.9583

CITATION

Hoi, L.M., Sun, Y., Lin, M., et al., 2025. Design of Intelligent Educational Mobile Apps with an Original Dataset for Chinese-Portuguese Translators. Forum for Linguistic Studies. 7(7): 696–718. DOI: https://doi.org/10.30564/fls.v7i7.9583

***Keywords:*** Automated Writing Evaluation; Chinese-Portuguese Translation Exercise Corpus; Common European Framework of Reference for Languages; fine-Tuning; Generative Artificial Intelligence; Large Language Models; Portuguese as a Foreign Language)

# 1. Introduction

In retrospect, numerous countries and regions, particularly in Asia, have been characterized by the governance of diverse ethnic languages, frequently resulting in multilingual official systems. Consequently, translation services have become a deeply embedded necessity for individuals residing or working within these areas. However, translation complexity increases significantly when the languages involved originate from disparate language families. For example, in the *Macao Special Administrative Region of China* (Macao), Chinese and Portuguese are used together; in Hong Kong, Chinese and English are used together; and in Singapore, English, Malay and Mandarin are used together.

Translation is a multifaceted discipline that incorporates diverse theoretical frameworks, methodologies, and analytical dimensions. Key considerations within translation include accuracy, fidelity, fluency, readability, acceptability, and usability[1]. Proficiency in translation is not attainable through expediency; it demands sustained effort and rigorous practice. Contemporary pedagogical approaches advocate for the application of *Self-determination Theory* (SDT) and *Mobile-assisted Language Learning* (MALL) to foster independent and intensive study, thus accelerating the acquisition process[2–4]. However, solitary practice proves to be ineffective without sufficient expert guidance. Consequently, there is a discernible absence of integrated automated evaluation tools within the existing mobile app targeted at Chinese-Portuguese translators. Therefore, a mobile app offering structured learning support for translation is a significant and currently unmet need.

This research details the development of a mobile app designed to facilitate the self-directed enhancement of translation skills. The mobile app comprises three core components: training materials, an automated evaluation system, and a user-friendly interface. Training materials, specifically the *Chinese-Portuguese Translation Exercise Corpus* (CPTEC), were developed by our research team and released publicly to contribute to language learning resources. Furthermore, this mobile app incorporates an *Analytic Hierarchy Process* (AHP) TQA model adapted from a previous research[1], enhanced with contemporary *machine learning* (ML) and LLM techniques for automated evaluation. Given the context of the study within a university in Macao, where students predominantly use Chinese as their native language and Portuguese as a *second language* (L2), the investigation specifically examines the proficiency of Chinese-Portuguese translation. The primary objective was to evaluate the efficacy of using the designed mobile app and associated learning materials in fostering student learning outcomes. To this end, we used the fine-tuning of LLMs to assess the impact of training content on translation performance. We anticipate that these findings will be beneficial to developers of educational mobile software applications.

The primary contributions of this article are summarized as follows:

- CPTEC release to the public to foster participation in the Chinese-Portuguese *Natural Language Processing* (NLP) and AI communities.
- Tailor make challenging exercises that can reinforce translation proficiency.
- Propose a novel hybrid grade system for automated translation evaluation.
- Illustrate the mechanism of computer-automated evaluation.
- Verify that the suggested exercises can strengthen translation skills by fine-tuning the LLM to show the improvement.

The organization of this article is as follows. Section 1 is an introduction. Section 2 discusses the background of LLM, MALL, SDT, TQA, and automated evaluation. Section 3 briefly explained the design of CPTEC. Section 4 is a methodology. The learning results are verified in Section 5, followed by future directions and conclusions in Section 6.

# 2. Literature Review

## 2.1. Self Learning through Mobile Apps

In contemporary society, mobile phones have become virtually indispensable tools, particularly within developed nations. Individuals increasingly rely on mobile apps for a wide range of activities, including purchasing goods, ordering food, and accessing transportation services. Recent research indicates that primary school students in Hong Kong spend a minimum of four hours a day on electronic devices (primarily mobile phones) during weekdays and six hours on weekends [5]. Although people often worry about the possible adverse effects of these devices, proponents argue that they help provide efficient and convenient learning opportunities [6, 7]. Consequently, contemporary educational theory advocates for the incorporation of mobile phones into independent learning frameworks. MALL research reveals that the use of portable devices can extend learning time and improve learning quality [3, 8]. Furthermore, the interactive nature and multimedia capabilities inherent in mobile apps contribute to a high degree of engagement of learners, often leading to sustained and intensive use.

On the other hand, mobile apps can leverage SDT to incorporate its core principles (autonomy, competence, and relatedness) and thereby meet fundamental psychological needs [2]. Recent studies have shown that users' propensity to use mobile apps has increased during the COVID-19 pandemic, a period during which traditional schooling has been disrupted [4]. Consequently, independent and intensive translation learning via mobile apps represents a potentially optimal pedagogical approach.

## 2.2. Translation Quality Assessment by Human

The evaluation of translations is inherently subjective and depends on factors such as time, geography, and context. For example, if a translation emphasizes strict fidelity to the original, it may sometimes produce interpretations that are difficult for people to understand and comprehend. Conversely, if a translation emphasizes clarifying the meaning of the original text while improving readability, it may inadvertently sacrifice the nuance and depth of the original. However, different audience groups will always prefer different translation methods.

Automated quantification of translation assessment items remains a significant challenge. The famous *China Accreditation Test for Translators and Interpreters* (CATTI) [9] examination is designed to certify different levels of professional translation competencies, but requires subjective evaluation by bilingual teachers and experts.

Throughout the years, researchers and language experts have proposed various evaluation systems to assess translation quality [10–12]. A recent study integrated various methodologies and introduced a novel TQA model structured around four dimensions [1]. The relationships between these criteria within the TQA model are illustrated in **Figure 1**.

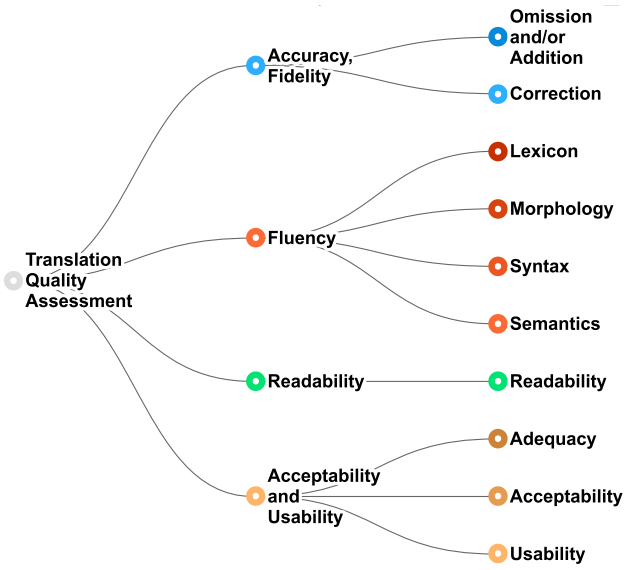


**Figure 1.** Analytic Hierarchy Process model of the Evaluation Index system of Chinese-Portuguese Machine Translation.

The primary contribution of this study lies in the determination of weight coefficients for each indicator within the Chinese-Portuguese translation evaluation index system [1]. This system facilitates a quantitative assessment of the relative importance of different dimensions of translation quality, thereby supporting the development of more sophisticated automated evaluation methodologies. The weight coefficients for each indicator are presented in **Table 1**.

**Table 1.** Weight coefficients for each indicator in the Chinese-Portuguese translation evaluation index system.

| First-Level | Weight | Second-Level | Weight | Overall |
|---|---|---|---|---|
| Accuracy, Fidelity | 0.525 | Omission a/o Addition | 0.162 | 8.50% |
| | | Correction | 0.838 | 44.00% |
| Fluency | 0.168 | Lexicon | 0.399 | 6.70% |
| | | Morphology | 0.148 | 2.50% |
| | | Syntax | 0.114 | 1.90% |
| | | Semantics | 0.339 | 5.70% |
| Readability | 0.229 | Readability | 1.000 | 22.90% |
| Acceptability and Usability | 0.078 | Adequacy | 0.149 | 1.16% |
| | | Acceptability | 0.393 | 3.07% |
| | | Utility | 0.458 | 3.57% |

## 2.3. Translation Quality Assessment by Machine

TQA is a demanding and resource-intensive process that requires a high degree of inter-rater reliability. As a result, the inherent complexity of TQA, especially under heavy workloads, has motivated the development of automated assessment methods to manage the large volumes of translations typically assessed in examinations and testing scenarios.

### 2.3.1. Automated Evaluation

There are multiple terms for automated assessment methods, including *automated writing evaluation* (AWE), *automated essay scoring* (AES), and *computerized essay scoring* (CES). Historically, efforts have been made to mitigate the challenges associated with manually evaluating written texts. Subsequently, software tools have been developed to identify and correct writing flaws. Typical examples of such systems include *Project Essay Grade* (PEG), *Intelligent Essay Assessor*™ (IEA), Electronic Essay Rater (e-rater®), and IntelliMetric. PEG uses statistical analysis, while IEA is based on LSA; in contrast, e-rater® and IntelliMetric are based on NLP[13]. However, these systems focus primarily on detecting spelling and grammatical errors in a specific language and lack the ability to perform detailed analysis of translation-related complexities, particularly semantics, acceptability, and usability. Consequently, learners experience limited learning progress and efficiency when using these automated assessment tools.

### 2.3.2. Machine Evaluation Standards

The annual North American Machine Translation Conference, hosted by the *North American Chapter of the Association for Computational Linguistics* (NAACL), attracts researchers and practitioners from academic institutions, research laboratories and well-known technology companies to attend the *workshop on machine translation* (WMT). WMT covers competitions that evaluate various aspects of machine translation, using a range of different scoring criteria, including the *BiLingual Evaluation Understudy* (BLEU), chrF, BERTScore, COMET, BLEURT, Prism and YiSi-1[14]. It is worth noting that BLEU is still the industry's widely recognized and mature benchmark for machine translation evaluation.

BLEU attempts to evaluate the quality of machine-generated text with the ground truth (reference translated text) by comparing the sentence similarities using n-gram overlap. It compares the n-grams of the translation to be evaluated and the reference translation and calculates the number of matching segments; the higher the number of matching segments, the better the quality of the translation to be evaluated[15].

Equation (1) represents the algorithm $f_{BLEU}$ for calculating the BLEU value. In addition, $BP$ (Brevity Penalty) is a point deduction mechanism for translations that are too short, $P_n$ is the score of each gram, and $w_n$ is the weight of each gram.

$$f_{BLEU} = BP \times \exp(\sum_{n=1}^{N} w_n \log P_n) \tag{1}$$

The *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) represents a scoring metric analogous to BLEU and is frequently used in the evaluation of the text summarization[16]. BLEU remains the prevailing standard utilized in international competitions, even numerous alternative metrics exist over the years. However, BLEU's reliance

on a single reference translation limits its effectiveness, as it may not accurately assess translations of comparable quality. Furthermore, BLEU provides limited insight into the specific nuances of translation errors, thereby hindering opportunities for performance improvement.

## 2.4. Large Language Models

Recently, researchers have increasingly used established NLP techniques to address text processing challenges. The rise of neural networks and AI technologies in the 2010s has significantly enhanced the capabilities of automated text processing, yielding results that are both remarkably effective and surprising.

In 2018, OpenAI released the *Generative Pre-trained Transformer* (GPT) language model, which garnered significant attention for its predictive capabilities [17–20]. Subsequently, in 2022, OpenAI introduced ChatGPT, leveraging the GPT-3 language model for text generation, achieving remarkable results [21]. With ongoing advancements, notably the transition to GPT-4 [22, 23], this LLM approach to language processing is increasingly dominating the market. The capability of ChatGPT to respond to queries in a manner that simulates human interaction has led to its designation as *Generative Artificial Intelligence* (GenAI), and its prominence within the market has subsequently flourished.

Generally speaking, LLMs need the acquisition and processing of substantial datasets and parameters, demand significant computational resources and, consequently, limit accessibility. Recent research has comparatively evaluated the capabilities and performance of prominent GenAI models currently available, including BLOOM, ChatGPT, Copilot, Gemini, Granite, LLaMA, Snowflake, and StarCoder [24, 25]. Beyond text generation, LLMs demonstrate versatility in handling multimedia content and executing a wide range of NLP tasks. Furthermore, the process of fine-tuning enables adjustments to model outputs without necessitating costly retraining infrastructure.

### 2.4.1. LLMs for Different Tasks

LLMs frequently incorporate different post-training methodologies, including supervised learning, few-shot learning, and domain-specific, alongside various fine-tuning styles such as transfer learning, instruction tuning, and alignment tuning, to achieve targeted task performance or facilitate prompt engineering. The enhanced capabilities and sophistication of LLM technology, relative to just a few years ago, now process tasks accurately and seamlessly including sentiment analysis, question answering, and document summarization. Consequently, the applications of LLM extend beyond academic research and development to encompass practical solutions within the business sector. Below is a summary of commonly used LLMs that solve various NLP challenges [26, 27].

- **Question Answering Model:** It refers to the task of extracting answers to questions from specified text, suitable for searching for answers in documents.
- **Sentence Similarity Model:** It refers to the task of calculating how similar the target sentences are to the source.
- **Summarization Model:** It refers to the task of generating a shorter version of a document while retaining its important information. Some models can extract text from raw input, while others can generate entirely new text.
- **Text Classification Model:** It refers to the task of assigning labels or categories to a given text. Some use cases include sentiment analysis, natural language inference, and assessing grammatical correctness.
- **Text-generation Model:** It refers to the task of generating a new text based on another text. For example, these models can fill in incomplete text or paraphrases.
- **Token Classification Model:** It refers to the task of understanding natural language, where some markers in a text are assigned a label. Some popular token classification subtasks are *Named Entity Recognition* (NER) and *Part-of-Speech* (PoS) tagging.
- **Translation Model:** It refers to the task of converting text from one language to another.
- **Zero-shot Classification Model:** It refers to the task of predicting categories that the model has not seen during training.

### 2.4.2. Fine-Tuning

Fine-tuning represents a highly effective strategy for optimizing LLMs to perform specific tasks. Two prevalent fine-tuning methodologies, conspicuously, *Low-Rank Adaptation* (LoRa) and *Quantized Low-Rank Adaptation* (QLoRa), are frequently encountered in academic papers regarding model optimization.

LoRa is a technique that updates a limited set of parameters, thereby generating a reduced subset of the base model for fine-tuning, rather than appending a new layer. As a result, this constrained update process facilitates accelerated fine-tuning because only a fraction of the base model requires modification [28]. QLoRa compresses the pre-trained model into a 4-bit representation. This quantization method reduces the overall size of the model while preserving a significant portion of the accuracy of the original weights [29].

# 3. Chinese-Portuguese Translation Exercise Corpus

*"Practice makes perfect"*. Consistent practice is demonstrably effective in fostering skill development. Consequently, we have developed a CPTEC dataset designed to facilitate practice for translators.

## 3.1. Design of Translation Exercise Corpus

We established a corpus team consisting of two bilingual teachers and nine translation students, three of whom were Chinese and six were Portuguese.

### 3.1.1. Design of the CPTEC Difficulties

The *Common European Framework of Reference for Languages* (CEFR) serves as an internationally recognized standard to describe the level of proficiency of language learners. Using a six-level scale, ranging from "*A*" (basic user) to "*C*" (proficient user), as shown in **Figure 2**, the CEFR facilitates the comparative and assessment analysis of language skills through different languages and establishes a unified framework for language education and evaluation [30].

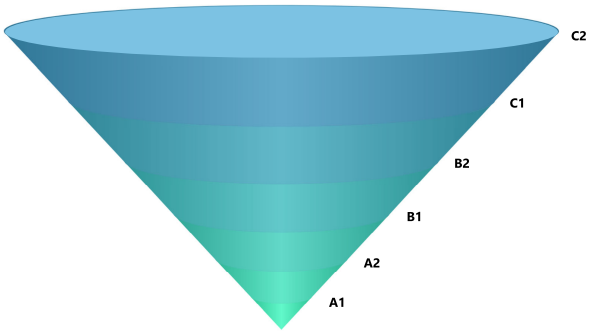


**Figure 2.** Common European Framework of Reference for Languages.

As a result, CEFR provided a suitable benchmark for establishing proficiency levels within CPTEC. The three levels of CEFR (A, B, and C) can be directly aligned with the CPTEC classifications of elementary, intermediate, and advanced. Detailed information on each CEFR level is presented as follows.

A1. **Lower-Beginner:** Learners are able to understand and use familiar everyday expressions and very basic phrases to meet specific types of needs.

A2. **Upper-Beginner:** Learners can understand sentences and common expressions relating to the areas of most immediate relevance (e.g., very basic personal and family information, shopping, employment).

B1. **Lower-Intermediate:** Learners are able to understand the main points of clear, standard input on familiar matter that is frequently encountered in work, school, leisure, etc.

B2. **Upper-Intermediate:** Learners can understand the main ideas of complex texts on both concrete and abstract topics, including technical discussions in their field of expertise.

C1. **Advanced:** Learners can comprehend a variety of demanding and lengthy texts and identify implicit meanings.

C2. **Mastery:** Learners can easily understand almost anything they hear or read.

### 3.1.2. CPTEC Schema

The CPTEC comprises nine distinct elements or data fields, including an identification number, the original Chinese text, reference translations, alternative vocabularies, and genre. Furthermore, each reference translation and alternative vocabulary is presented at three hierarchical levels: elementary, intermediate, and advanced. The CPTEC schema is shown in **Table 2**.

The proposed schema design aims to accommodate three distinct levels of complexity in translating Chinese sentences into Portuguese. Subsequently, two bilingual teachers delineated twelve categories (art, economic, education, environmental science, festival, food culture, literacy, medical science, politics, sport, technology and tourism) to serve as the basis for student-generated content within each corpus.

**Table 2.** The schema of CPTEC.

| | Field Name | Data Type | Description |
|---|---|---|---|
| 1 | ID | Integer | identification Number |
| 2 | ZH | String | original text in Chinese |
| 3 | RT_A | String | elementary reference translation |
| 4 | RT_B | String | intermediate reference translation |
| 5 | RT_C | String | advanced reference translation |
| 6 | AV_A | String | elementary alternative vocabulary |
| 7 | AV_B | String | intermediate alternative vocabulary |
| 8 | AV_C | String | advanced alternative vocabulary |
| 9 | GENRE | String | domain of the original text |

### 3.1.3. Core Translation Concepts

To facilitate the development of translation skills among users, the corpus prioritizes the application of core translation concepts over the generation of isolated sentences. Consequently, two bilingual teachers claimed five fundamental principles (fidelity, fluency, functionality, practicality, and acceptability) for students to consider during the translation process.

Some of the core concepts are translation is not just a conversion of the form of language, but is more on the transmission of the function and meaning of the source language text. A good translation should be natural and smooth in form. It always conforms to the expression habits of the target language [31]. Translation should not only pay attention to the form of language, but also consider context and communicative intention. A good translation should correctly convey the pragmatic meaning of the original text, including implicit meaning, politeness, humor, etc. [32]. Translation is a complex social behavior affected by many factors [33].

The construction of the corpus necessitated several specific requirements. Three Chinese students generated the initial set of original Chinese sentences (ZH) comprising approximately 500 sentences per category. These sentences were designed to be syntactically complete, with a character count ranging from 15 to 30, and incorporating both noun and verb elements – a constraint deemed essential for facilitating the subsequent creation of distinct Portuguese translation levels. Subsequently, six Portuguese students selected two categories for translation, tasked with producing three varying levels of Portuguese sentences (RT_A, RT_B and RT_C) alongside alternative vocabulary choices (AV_A, AV_B and AV_C). Finally, two bilingual teachers oversaw the entire process, rigorously verifying the work to eliminate redundancy and guaranteeing the presence of three distinct difficulty levels within the corpus. Consequently, the corpus team successfully compiled a collection of approximately 5,000 corpora.

### 3.1.4. The Potential of CPTEC

CEFR is designed for different applications that encompass several key areas [30]. Given that CPTEC aligns with CEFR, it can be utilized to develop a variety of educational systems and mobile apps aimed at student training.

- **Language learning and teaching:** It provides a structured approach to language learning and teaching, to set goals, to measure progress, and to identify areas for improvement.
- **Language assessment:** Language assessment/certification entities use to develop exams that measure different levels of language proficiency.
- **International mobility:** Employers, educational institutions, and immigration authorities use it to assess the language proficiency of people who want to work, study, or live abroad.
- **Curriculum development:** It provides a framework to consolidate language proficiency that aligns with the goals of the learner.
- **Artificial Intelligence:** Use LLM and fine-tuning techniques to solve specific NLP problems.

## 3.2. Open-Source Dataset

CPTEC offers significant potential for both software development and advancements within the field of education. Furthermore, its utilization in fine-tuning LLMs facilitates the creation of numerous AI models capable of addressing complex NLP challenges. As a consequence, CPTEC's ca-

pabilities represent a substantial and valuable resource.

After carefully reviewing every word in CPTEC, we officially release it to the public [34] as shown in **Figure 3**, hoping to benefit the field of Chinese-Portuguese NLP.

Apache Parquet is a popular data format for processing complex datasets within column-oriented databases [35]. Most AI community platforms facilitate data visualization through web interfaces, as illustrated in **Figure 3**, and support conversion to a variety of formats, including Apache Parquet, CSV, and JSON. Commonly employed formats such as JSON, frequently utilized for data exchange between networked servers, are detailed below.

```
{
  "ID" : "sequential ID number",
  "ZH" : "Original sentence in Chinese",
  "RT_A" : "A1−A2 level reference translation",
  "RT_B" : "B1−B2 level reference translation",
  "RT_C" : "C1−C2 level reference translation",
  "AV_A" : "A1−A2 level alternative vocabulary",
  "AV_B" : "B1−B2 level alternative vocabulary",
  "AV_C" : "C1−C2 level alternative vocabulary",
  "GENRE" : "domain of a sentence",
},
{ · · · }
```

Sample Sentences and Verbs

| basic_sent<br>string | independent_sent<br>string | proficient_sent<br>string | basic_verb<br>string | independent_verb<br>string | proficient_verb<br>string | category<br>string |
|---|---|---|---|---|---|---|
| "O Sonho do Pavilhão Vermelho" é um livro com muitas fantasias e… | Com as suas imagens ricas e metáforas profundas "O Sonho do… | Por meio de ilustrações magníficas e recursos expressivos profundos, o… | imaginação, famoso, pretender, debater (sobre, com, em), analisar… | marcantes, essência, sentido, actualmente, presentemente, abordar… | figuras de estilo, esplêndido, retratos, fundamento, examinar (com… | literature |
| O "Romance dos Três Reino" conta histórias verdadeiras do passado a… | O clássico "Romance dos Três Reinos" retrata o clima e a luta social de… | Tendo como pano de fundo acontecimentos históricos reais, o… | reais, casos, situações, episódios, época | verídicas, descrever (em, sobre, com), ambiente, confronto,… | contexto, reproduzir (em, com), período, minuciosamente, criar (em,… | literature |
| A literatura é uma ponte para a inteligência e o sentimento humano… | A literatura é uma ponte para o mundo da sabedoria e das emoções… | A literatura é um vínculo para o mundo do intelecto e das emoções… | ligação, conexão, entender (de), captar, puder | domínio, conhecimento, sentimentos, dar, proporcionar | elo, oferecer, essência, fundamento, possibilitar | literature |
| Era uma vez, numa pequena aldeia, um rapaz normal muda de vida por… | A história começa numa pequena aldeia onde a vida de um jovem comu… | O enredo inicia-se numa pequena aldeia onde o destino de um… | alterar, devido a, localidade, menino, garoto | conto, manuscrito, povoado, futuro, rumo | narrativa, vilarejo, por motivo de, fado, na qual, enigmático | literature |
| Estou no topo da montanha, unido com a natureza, a sentir o bater d… | Estou no pico da montanha, juntamente com a natureza, sentindo… | Estou no cume da montanha, em sintonia com a natureza, sentindo o… | cimo, alto, solo, ruído, zumbido | encontrar (com), conectar (com, em), meio ambiente, solo, superfície | harmonia, vivenciar, brisa, palpitar, murmúrio | literature |
| O tempo é como um grande rio que não para de correr durante os anos. | O tempo parece um longo rio que corre interminavelmente ao longo do… | O tempo lembra um rio fluente que atravessa o dilúvio dos anos. | contínua, fluído, duração, curso de água, movimentar (em, sobre) | assemelhar (com), extenso, fluir, infinitamente, através do tempo | dar a sensação (de), recordar (de, sobre), percorrer, transpor, décadas | literature |
| A história principal é sobre amor proibido, com muito amor, luta e… | O eixo central da história fala sobre amor proibido, com muita… | A espinha dorsal da história é um amor proibido, cheio de paixão,… | foco da história, romance, relacionamento, sofrimento, infinito | acerca de, a respeito de, descrever (em, com, sobre), interminável,… | amor clandestino, não correspondido, amor intenso, paixão ardente,… | literature |
| Novos medicamentos trazem esperanças, pois podem tratar o… | Novos medicamentos demonstraram ser eficazes no tratamento do cancro,… | Novas esperanças crescem à medida que novos medicamentos se revelam… | porque, curar (com), remédios, recentes, boas expectativas | fármacos, comprovar (com, sobre, em), provar (com, sobre, em),… | recente, aumentar, enquanto, demonstrar (em, com), apresentar… | medical science |
| O grupo de médicos usa tecnologia nova para fazer cirurgias correcta… | A equipa médica utiliza tecnologia moderna para realizar cirurgias… | A equipa médica utiliza tecnologia avançada para efectuar cirurgias… | técnica, meio, tratamentos, operações, reduzir | manusear (em, com), recurso, métodos, a fim de, com o objectivo… | com a finalidade de, com intenção de, executar (com), rigorosas,… | medical science |
| Os médicos por trabalharem há muito em tempo doenças raras, fizeram… | Há muito que os médicos estão empenhados na investigação de… | O empenho de longa data dos médicos na investigação das doenças raras… | por causa de, desde há muito tempo, incomuns, levar (com, sobre), deste… | achados, significativas, relevantes, valorosas, curas | dedicação, afinco, profissionais de saúde, patologias, encaminhou | medical science |

**Figure 3.** Samples of the CPTEC dataset.

# 4. Methodology

This article investigates the potential of incorporating our training content, a graded assessment system, and automated evaluation mechanisms to leverage translation proficiency through a mobile app. Specifically, this section outlines the procedural implementation of the mobile app utilizing our established training materials. Furthermore, it details our novel grading system to detect individual learner weaknesses.

## 4.1. The Workflow of a Translation Exercise

Educational mobile apps can include many innovative features and interesting interactions to enhance the user experience. This section focuses on demonstrating a translation exercise using CPTEC and how this exercise can reinforce translation skills.

Educational mobile apps frequently incorporate innovative features and interactive elements designed to enhance the user experience. This section presents examples of translation exercises that use CPTEC to illustrate its potential to enhance translation proficiency.

**Figure 4** shows a storyboard wireframe of the translation exercise in the mobile app. Initially, the user is requested to log in and select a proficiency level: beginner (1), intermediate (2), or advanced (3). Following this selection, the mobile app retrieves a record from the CPTEC dataset and generates a translation task consisting of a Chinese sentence paired with suggested vocabulary. Upon completion of the translation, the application automatically assesses the user's response and displays a final score.

## 4.2. Apps Development Environment

Previous research indicated that our target audience predominantly used Android and iOS mobile devices for Portuguese language learning [4]. Therefore, resource allocation would be most effectively directed towards the development

of mobile apps specifically for these platforms, rather than supporting a broad range of electronic devices.

Conventionally, application development for specific platforms relied on native code, such as Swift for iOS and Kotlin for Android, resulting in the maintenance of disparate codebases and development environments. For applications that do not necessitate extensive utilization of hardware modules or native device functionalities, a cross-platform framework offers an adequate solution for development requirements.

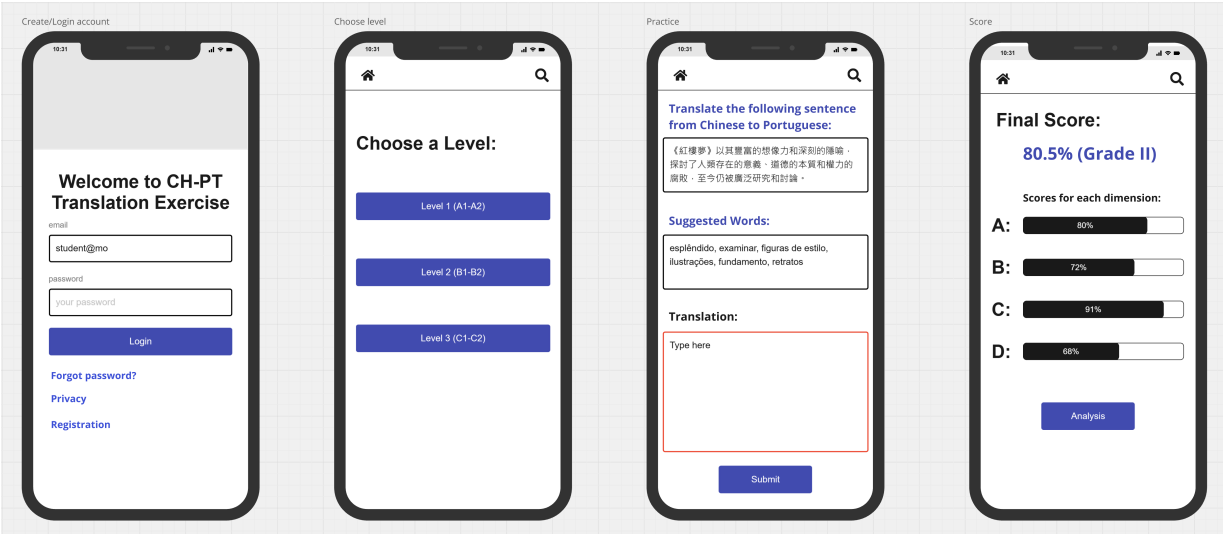


**Figure 4.** The storyboard wireframe for the translation exercise.

One of the beauty of writing cross-platform code is that it allows compiling the same code for different electronic device platforms (Android and iOS). Some cross-platform frameworks can be found in the market, including Meta's React Native, Google's Flutter, Microsoft's Xamarin, etc. In this article, React Native is selected as the mobile app development framework with JSX markup as the programming language. Consequently, we can focus on how to promote translation skills through mobile apps instead of debugging tricky native code for the diversifying platforms.

The inherent advantage of cross-platform code development is the ability to compile the same code base for different electronic platforms, such as Android and iOS. There are currently several cross-platform frameworks available, including Meta's React Native, Google's Flutter, and Microsoft's Xamarin. This article will rely on React Native as a mobile app development framework and uses JSX markup as its programming language. Therefore, this article will explore strategies for developing translation skills through mobile app development, rather than addressing the complexity of debugging native code across platforms.

## 4.3. Design of the Translation Exercises

Our mobile app designers were able to produce a multitude of rich storyboards. Since we cannot illustrate them all, this section aims at comprehensively discussing our approach to reinforce translation skills through the app.

### 4.3.1. Translation Exercises Creation

The translation exercises begin with choosing different levels (elementary, intermediate, and advanced) and types (domains of sentences) by users. Then, it randomly picks a CPTEC record that consists of nine fields, as in **Figure 3** with the selected genre, and generates an exercise page for users. This page uses the Chinese sentence as the question, and it requires users to translate the Chinese text into Portuguese with the suggested words. Moreover, each exercise will randomly pick some words from the reference translation together with the alternative vocabularies of the corresponding level as a set of suggested words.

The translation exercises begin with user selection of proficiency level (elementary, intermediate, or advanced) and a sentence domain type. Subsequently, the system randomly

retrieves a CPTEC record, as shown in **Figure 3**, comprising nine fields, and generates a dedicated exercise page. This page uses Chinese sentences as questions, asking users to translate Chinese text into Portuguese using the suggested vocabularies provided. Moreover, the system randomly incorporates a selection of words from the reference translation alongside alternative vocabulary options appropriate for the user's chosen proficiency level.

Equation (1) outlines the algorithm $f_s$ for retrieving an exercise with the input parameters $l$ and $g$ from CPTEC. Specifically, $l$ denotes level, $g$ denotes genre, $s$ signifies a CPTEC record, and $S_{CPTEC}$ represents the corresponding CPTEC dataset.

$$f_s(l, g) = \{s_{lg} | s \in S_{CPTEC}\} \tag{1}$$

With the aim of increasing the challenge for users, the suggested words do not use simple and daily words. These words encompass articles (a, as, o, os, etc.), conjunctions (como, e, mais, mesmo, ou, perque, porque, etc.), determiners (aquele, esta, este, qual, todo, etc.), numbers (uma, dois, três, etc.), prepositions (até, com, de, em, entre, para, etc.) and pronouns (ella, elle, eu, nos, tu, vos, etc.)[36, 37]. As a result, an exception list was established to encompass the aforementioned terms. This approach aimed to provide users with more sophisticated noun and verb alternatives within the reference translation text, facilitating targeted practice. According to core translation concepts mentioned in Section 3.1.3, we anticipate that this will facilitate the development of a more nuanced vocabulary, enabling a more precise articulation of the meaning of the original text.

The exception list comprises approximately 500 words, predominantly consisting of shorter terms (fewer than six characters). A statistical analysis of this list is presented in **Table 3**. Therefore, a more efficient strategy is to filter the list to retain only words longer than six characters, thus eliminating the need for a full comprehensive enumeration.

Equation (2) represents an algorithm for picking suggested words from the reference translation. The algorithm $f_{rt}$ takes two input parameters (reference translation $s_{rt}$ and number of words $n$) with the word $w$ from the reference translation, not in the exception list $E$, and the length of the word $w$ is greater than or equal to six.

$$\begin{aligned} f_{rt}(s_{rt}, n) = \{w_1, w_2, ...w_n | \\ w \in s_{rt}, w \notin E, len(w) \geq 6\} \end{aligned} \tag{2}$$

**Table 3.** Statistics for the list of exceptions.

| Length [1] | Count | Length [1] | Count |
|---|---|---|---|
| 1 | 9 | 9 | 13 |
| 2 | **64** | 10 | 12 |
| 3 | **94** | 11 | 7 |
| 4 | **102** | 12 | 5 |
| 5 | **77** | 13 | 2 |
| 6 | 32 | 14 | 1 |
| 7 | 22 | 15 | 1 |
| 8 | 15 | 16 | 1 |

[1] Length of the exception words.

Equation (3) represents an algorithm for picking suggested words from the alternative vocabularies. The algorithm $f_{av}$ takes two input parameters (alternative vocabulary $s_{av}$ and number of words $n$), and this algorithm picks only $n$ alternative vocabularies.

$$f_{av}(s_{av}, n) = \{w_1, w_2, ...w_n | w \in s_{av}\} \tag{3}$$

Equation (4) represents the algorithm for picking the suggested words to form an exercise. The algorithm $f_{sw}$ takes two input parameters (a CPTEC record with a particular level and genre $s_{lg}$ and the number of words $n$). Then, it calls two algorithms $f_{av}$ and $f_{rt}$ with correspondingly $s_{av}$ and $s_{rt}$, and a parameter $n$. Notice that $s_{av}$ and $s_{rt}$ belong to the record of $s_{lg}$.

$$\begin{aligned} f_{sw}(s_{lg}, n) = f_{av}(s_{av}, \frac{n}{2}) \cup f_{rt}(s_{rt}, \frac{n}{2}) \\ where \quad \{s_{av}, s_{rt}\} \in s_{lg} \end{aligned} \tag{4}$$

The upper equation facilitates the retrieval of pertinent information from CPTEC for mobile software developers undertaking translation exercises.

### 4.3.2. Translation Exercises Analysis

There are many ways to evaluate the effectiveness of these exercises in improving translation skills. Traditional evaluations often rely on questionnaires, a process that requires extensive user feedback. Alternatively, one could mimic the approach taken in previous studies and collect data to perform rigorous analysis directly from mobile apps[4]. Furthermore, these exercises can also be validated using contemporary LLM techniques, which will be detailed in Section 5.

## 4.4. Design of the Automated Evaluation System

Upon completion of the translation exercise, the mobile app should be able to automatically identify errors in the user's feedback and provide targeted recommendations. As a result, users can use the mobile app to effectively discover the weak that is associated with translation. This section details a new evaluation system for evaluating translated texts.

The proposed automatic assessment system should incorporate both the criteria evaluated through manual assessment, as detailed in Section 2.2, and the industry standards outlined in Section 2.3. Therefore, it implements the TQA system presented in **Table 1** as an agent of human evaluation, while employing the BLEU score as a measure of machine evaluation.

### 4.4.1. Hybrid Grade System

The widespread availability and user-friendliness of smartphones often leads to learning the same content repeatedly, especially when faced with difficult questions. As a result, users may inadvertently memorize the standard answers (reference translations) provided, in which case full marks should be given.

In consequence, the evaluation results will be determined by the BLEU value calculated from the reference translation. A BLEU value of 0.9 or higher will result in a first-level rating. On the other hand, scores below 0.9 will require the application of a new TQA system to determine the final score.

To facilitate automated evaluation of translated texts within the mobile app, we implemented a hybrid grade system with a five-level scale, as detailed in **Table 4**.

**Table 4.** A 5-level Grade System.

| Grade | Score | Remark |
|---|---|---|
| I. | ≥90% | Excellent |
| II. | 70 ∼89% | Good |
| III. | 50 ∼69% | Fair |
| IV. | 20 ∼49% | Fail |
| V. | ≤19% | Not even close |

### 4.4.2. Novel Translation Quality Assessment System

TQA has historically been centered on human evaluation methodologies, especially translation evaluation frameworks such as human-based quality assurance models (shown in **Figure 1**). These frameworks prioritize language fidelity, fluency, and usability, providing comprehensive evaluation criteria that include accuracy, readability, and acceptability. Human-based assessments, while insightful, are limited by subjectivity, inconsistency, and reliance on individual expertise[38]. Moreover, they lack scalability and efficiency, as large-scale and high-speed machine translations require automated evaluation solutions[39, 40].

Given the existing landscape of automatic translation evaluation systems, this approach leverages a model illustrated in **Figure 1**, refining its established criteria and adapting its structural framework for implementation within a computer-driven evaluation system. This transition to a fully automated framework facilitates systematic modifications to the weighting and application of individual evaluation elements, leading to a more scalable and automated assessment process.

Leveraging established frameworks and the foundational translation evaluation concepts previously outlined, we developed a machine-based approach to convert salient linguistic and contextual features into quantifiable metrics. These metrics are detailed and justified in **Table 5**.

**Table 5.** The weights of the novel Translation Quality Assessment System.

| First-Level | Second-Level | Weight |
|---|---|---|
| A. Accuracy, Fidelity | a1. Lexicon | 44.0% |
| | a2. Correction | 15.2% |
| B. Fluency | b1. Morphology | 2.5% |
| | b2. Syntax | 1.9% |
| C. Readability | c1. Readability | 22.9% |
| D. Acceptability, Usability | d1. Semantics | 9.3% |
| | d2. Adequacy | 4.2% |

a1. **Lexicon:** Adherence to mandatory or recommended terminology ensures domain-specific accuracy. This is particularly critical for technical translations, where incorrect terminology can lead to misunderstandings or errors in context-sensitive fields such as medicine or law. Effective lexical evaluation can significantly improve domain-specific translation accuracy[41, 42].

a2. **Correction:** Precision is prioritized by focusing on the accurate representation of critical factual details such as dates, names, and numerical data. This is essential for maintaining the integrity of translations, especially in

formal or legal documents, where such details have significant implications. Ensuring corrections minimizes factual discrepancies.

b1. **Morphology:** This metric evaluates the proper use of word forms, such as prefixes, suffixes, tense, pluralization, and case endings, ensuring grammatical correctness within the translated text. Using the proper morphological alignment ensures that words convey the intended meaning and grammatical relationship in a sentence, and clarity improvement [43].

b2. **Syntax:** Syntax is very effective in *neural machine translation* (NMT), which can assess the structural organization of sentences, including word order, clause structure, and grammatical dependencies [44]. This metric improves coherence and reduces ambiguity by ensuring syntactic accuracy, making sure sentences conform to the grammatical rules of the target language. This is especially important for languages with rigid word order, where misplaced elements can obscure meaning or alter intent.

c1. **Readability:** Ensuring the readability involves identifying irrelevant or excessive characters, maintaining appropriate sentence length, and simplifying overly complex constructions. Readability is a critical factor in user satisfaction, as poorly structured text often leads to frustration or misinterpretation. Readability metrics improve accessibility, particularly for general audiences [45].

d1. **Semantics:** Semantic evaluation focuses on the meaning conveyed by the translation to ensure that it is consistent with the original text. This involves evaluating whether the translated text accurately captures the meaning of nuances, idiomatic expressions, and specific contexts. Semantic fidelity in interactions has been shown to be effective on the NMT related tasks. It ensures that the translation retains the intended information and avoids misleading interpretations [46, 47].

d2. **Adequacy:** Adequacy measures how comprehensively the translated text reflects the content of the original text. Semantics assesses meaning, while sufficiency ensures that the information is complete and containing all necessary details [48, 49].

Subsequently, if the BLEU score falls below 90%, the TQA system (**Table 5**) can be adopted to quantitatively assess the final performance. To facilitate a more intuitive understanding of the relative importance of each dimension, **Table 5** is visualized as a two-layered pie chart, as shown in **Figure 5**.


**Figure 5.** Visualizing the scoring system of TQA.

The proposed TQA system, detailed in this section, aims to facilitate machine-based evaluations that emulate human judgment. Currently, language learners are limited to practicing reference translations and tend to copy model answers, which hinders the generation of original, potentially high-quality translated texts.

## 4.5. Implementation of Computer Automated Evaluation

Given the integration of a quantitative TQA marking scheme (detailed in Section 4.4.2) within our proposed hybrid grade system (described in Section 4.4.1), the human-like scoring process is rendered less abstract. Consequently, automated evaluation of translation quality within a mobile app becomes feasible. **Figure 6** illustrates the automated translation evaluation workflow. This section details the computational methods adopted to determine each dimension outlined in **Table 5**, leveraging contemporary technologies. In addition, it outlines strategies for automatically calculating each dimension using modern techniques.

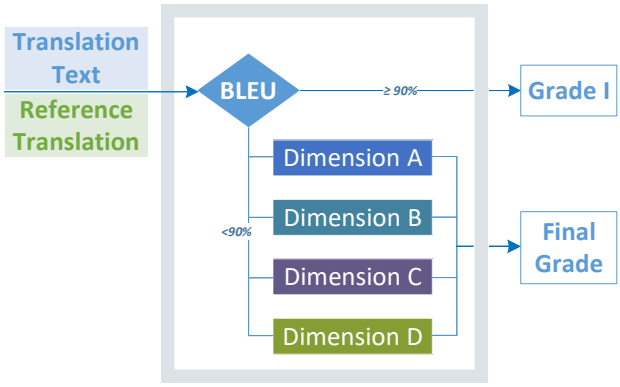


**Figure 6.** The workflow of automated translation evaluation.

### 4.5.1. Used of Words

**Table 5** indicates that dimension "A" exhibits the highest relative percentage compared to other dimensions. This dimension primarily assesses the appropriate use and accuracy of vocabulary in translation contexts. Given that suggested vocabulary is provided within the exercises, users are expected to employ these terms when constructing their translated texts. Hence, the mobile app can evaluate the user's ability to effectively integrate these terms correctly for scoring purposes.

The mobile app can verify the accuracy of the translation by carefully examining subtle linguistic elements that include dates, times, and monetary values. For example, different conventions for currency separators between Chinese and Portuguese posed significant challenges. Specifically, the Chinese representation of $1,234,567.89 is equivalent to $1.234.567,89 in Portuguese. Consequently, programming languages such as Python incorporate native libraries, often referred to as locales[50], which facilitate the conversion of currency representations across different national standards. These libraries allow you to validate the correct format of currency amounts.

Moreover, formal writing employs a unique system for representing numbers. Numeric characters, such as 1, 2, and 3, can be rendered in Portuguese as (um, dois, três), ordinal numbers (primeiro, segundo, terceiro, etc.), months (Janeiro, Fevereiro, Marchar, etc.), and days of the week (Segunda-feira, Terça-feira, Quarta-feira, etc.). To facilitate verification of these textual representations in Chinese and Portuguese, the following dictionary has been developed.

```
{
  "name": "number",
  "pt": [{ "1": "um" },
    { "2": "dois" }, { "3": "três" },
    { ··· }],
  "zh": [{ ··· }],
},{
  "name": "ordinal_number",
  "pt": [{ "1": "primeiro" },
    { "2": "segundo" }, { "3": "terceiro" },
    { ··· }],
  "zh": [{ ··· }],
},{
  "name": "month",
  "pt": [{ "1": "Janeiro" },
    { "2": "Fevereiro" }, { "3": "Marchar" },
    { ··· }],
  "zh": [{ ··· }],
}, {
...
}
```

### 4.5.2. Grammar Checking

In **Table 5**, the dimension "B" focuses its analysis on the syntactic elements of sentences, specifically examining aspects such as punctuation, capitalization, spelling check, and grammatical correctness. Several *Free and Open Source Software* (FOSS) tools are available to verify Portuguese sentences, including LanguageTool, QuillBot, and Writefull. Additionally, numerous JavaScript libraries available on GitHub offer potential integration within development projects.

Hence, the mobile app can integrate a grammar checking library to validate the syntactic correctness of the translation text. A reduction in the final score will be applied to this dimension should the grammar checker identify any suggested revisions.

### 4.5.3. Format and Ratio

The dimension "C" in **Table 5** concerns the readability of the translated text. Recent advancements in AI, particularly the proliferation of generative AI and machine translation technologies, have led to a common practice of utilizing these tools to produce initial drafts of translations, subsequently followed by manual revisions.

Upon utilizing AI tools, residual artifacts are readily introduced if meticulous review and correction are not undertaken. Specifically, these tools may employ less frequently encountered vocabulary. Furthermore, their construction relies on expansive datasets which can inadvertently include extraneous textual elements, such as foreign language terminology (e.g., Spanish substituting for Portuguese), non-printing control characters, corrupted text, superfluous whitespace, and erroneous characters. Hence, programs designed to identify and mitigate these anomalies are warranted.

An easily observable indicator of the use of these AI tools is the ratio of the length of the original and translated text. Generative AI models, when encountering unfamiliar content, often augment the original material with superfluous information, whereas machine translation systems frequently exhibit a reduction in information, effectively omitting key elements. Hence, a program could be developed to quantify the proportion of Chinese sentences successfully translated into Portuguese sentences.

In prior research [51], a statistical analysis of a corpus of approximately 10 million Chinese-Portuguese parallel texts sourced from various news websites is summarized in **Table 6**. The results indicate an average sentence length ratio of 1.37, a median ratio of 1.33, and a standard deviation of 2.04. Subsequently, we plotted the length ratio distribution as in **Figure 7**, a bell-shaped graph is formed, showing that most of the distribution is in the middle of the graph.

**Table 6.** Statistics of a Chinese-Portuguese Parallel Corpus.

| Range [1] | Portion (%) | Range [1] | Portion (%) |
|---|---|---|---|
| <0.25 | 0.10 | 1.75 ~ 1.99 | **8.94** |
| 0.25 ~ 0.49 | 0.51 | 2.00 ~ 2.24 | **4.05** |
| 0.50 ~ 0.74 | 2.76 | 2.25 ~ 2.49 | 1.44 |
| 0.75 ~ 0.99 | **10.75** | 2.50 ~ 2.74 | 0.60 |
| 1.00 ~ 1.24 | **25.39** | 2.75 ~ 2.99 | 0.26 |
| 1.25 ~ 1.49 | **27.17** | ≥3.00 | 0.13 |
| 1.50 ~ 1.74 | **17.91** | | |

[1] Sentence length ratio of Chinese to Portuguese.

In short, given that corpora may exhibit significant variability in length, ranging from concise phrases and titles to extended paragraphs and multiple sentences, a suitable range for the Chinese-Portuguese sentence-length ratio is estimated to be between 0.75 and 2.25. Equation (5) demonstrates that the ratio of the source text $s$ to the target text $t$ falls within this approximate range.

$$f(s,t) = \frac{len(t)}{len(s)}, \quad where \quad 0.75 \leq f < 2.25 \qquad (5)$$

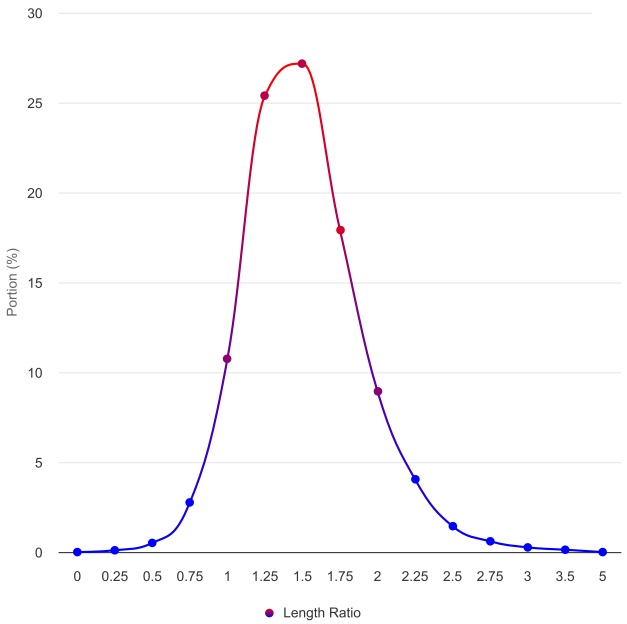


**Figure 7.** The distribution of the sentence length ratio.

### 4.5.4. Semantic Analysis

Last but not least, dimension "D" assesses the extent to which the translated text effectively conveys the semantics of the original text. With contemporary NLP and LLM technologies, it is possible to compare the meaning of the translation with that of a reference translation.

The development of a Portuguese LLM enables the extraction of key insights through a comparative analysis process. Specifically, the LLM can be prompted to articulate the central theme of two sentences, and subsequently, their relevance can be quantified using metrics such as cosine similarity or n-gram analysis. Furthermore, the LLM can be instructed to identify and delineate the underlying intentions (including nuance, tone, and implicit meanings) of the sentences, thus offering additional comparative data. Technically speaking, this iterative process of eliciting responses from the LLM via structured prompts is commonly referred to as prompt engineering.

**Algorithm 1** details the process of constructing an LLM capable of comparing a translated text with its corresponding reference translation. First, an LLM of type "text generation" or "text summarization" is selected, and the relevant Portuguese dataset is identified. Subsequently, the model and tokenizer are defined and initialized, and a pipeline is created that takes these components as input parameters. Finally, the input text $t$ and the prompt $p$ are fed into the pipeline to produce the final output.

**Algorithm 1** Programming Logic of using LLM

**Input:** $p, t$ ▷ p:prompt, t:text

$model_{pt} \leftarrow LLM_{pt}$
$model_{llm} \leftarrow Pretrained(model_{pt})$
$token \leftarrow Pretrained(model_{pt})$
$pipeline \leftarrow Pipeline(model_{llm}, token, type)$
▷ type: generation or summarization
**procedure** CALL_LLM
    $result \leftarrow pipeline(p, t)$
**return** $result$

This section details several mechanisms for simulating human evaluation. Provides users with valuable feedback to support practice and promote the generation of high-quality translations, rather than directing them toward a predetermined model response.

### 4.6. Intelligent Recommendation

Upon completion of data collection and analysis, AI can be implemented to address specific weaknesses. The previous sections detailed methodologies for evaluating translation quality and pinpointing areas for improvement. This section will subsequently describe the application of AI to mitigate these individual deficiencies.

Upon completion of the practice exercises through the mobile app, the resulting performance is visualized as depicted in the rightmost panel of **Figure 4**. Due to the dimensional scoring system, areas of weakness are readily identifiable. Furthermore, the application should subsequently offer targeted improvement strategies, including the provision of reference answers, pertinent learning resources, review of self-assessment, and supplementary exercises tailored to specific dimensions.

One of the challenges Chinese students face when learning Portuguese is often the accurate conjugation of verbs. Portuguese has more than 100 different verb conjugation patterns, each with about 70 conjugations [52]. Hence, mobile app developers can implement the function of showing the complete conjugation table of a specific verb in the Portuguese dictionary. These mobile apps can also provide teaching guidance by allowing users to search for examples of vocabulary misuse in the CPTEC. Using AI, such mobile apps can recommend tailored Portuguese learning materials based on the weaknesses of individual students.

In fact, there are numerous approaches to generate recommendations, and we encourage innovative mobile software designers and developers to implement novel solutions.

## 5. Verification

The main contribution of this article is the CPTEC dataset, which we expect will significantly improve the state of translation. Despite the fact that we have invested a lot of effort in perfecting the dataset, demonstrating its usefulness has proven challenging without empirical validation. As mentioned earlier, the implementation of user profiles is time-consuming and may require user consent. Using recent advances in the field of AI, we are now able to fine-tune LLMs for simulation. We can therefore train these LLMs using the CPTEC dataset to evaluate their translation capabilities at different levels of complexity. Specifically, evaluating the ability of LLMs to accurately translate Chinese sentences into three different Portuguese registers (elementary, intermediate, and advanced) will provide strong evidence for the effectiveness of the CPTEC dataset as supplementary training material.

The experimental methodology comprised the following steps: preparation of the dataset, selection of an appropriate LLM, training of the model with fine-tuning, and analysis of subsequent results.

### 5.1. Dataset Preparation

LLMs perform well on a wide range of NLP tasks; however, their architecture is inherently designed for general-purpose applications. Fine-tuning is used to specialize LLMs for specific tasks. This process requires a training dataset that contains instructions, inputs, and corresponding outputs. Therefore, the data in the CPTEC dataset must be defined and transformed to produce these necessary components.

#### 5.1.1. Instruction Definition

Specifying instructions presents a significant challenge and substantially impacts the output generated by LLMs. From a technical perspective, this process of defining instruction is often referred to as “Prompt Engineering”.

LLMs currently operate primarily as translation tools, necessitating precise and succinct instructions to minimize

extraneous information in their outputs. Consequently, it is crucial to build prompts with exceptional clarity and conciseness. For certain LLMs, incorporating specific phrases listed below within the instruction may be beneficial.

"*…you are an expert in…*"
"*…do not include anything other than the corresponding translation…*"
"*…do not duplicate translations…*"
"*…do not include other languages…*"

The instructions comprise a series of sentences and three variables: $V_{av}$, $V_{CEFR}$, and $V_{level}$. Following the structure outlined in the sentences stated below, these variables will be generated based on CPTEC, as detailed in **Table 7**. Consequently, each record within the CPTEC dataset can produce three distinct instructions.

*You are an expert in the field of Chinese-Portuguese translation. For any text I type next, you must complete a translation into $V_{level}$ Portuguese using vocabularies ($V_{av}$) or equivalent to the vocabularies of level $V_{CEFR}$ of the Common European Framework of Reference for Languages. The following is the Chinese text, please start translating:*

**Table 7.** The variables for generating the instructions, input, and output.

| Id | Instructions [1] | | | Input [1] | Output [1] |
|---|---|---|---|---|---|
| | $V_{level}$ | $V_{av}$ [2] | $V_{CEFR}$ | $V_{zh}$ [3] | $V_{rt}$ [4] |
| 1 | elementary | AV_A | A1-A2 | ZH | RT_A |
| 2 | intermediate | AV_B | B1-B2 | ZH | RT_B |
| 3 | advanced | AV_C | C1-C2 | ZH | RT_C |

[1] data fields from **Table 2**
[2] av = alternative vocabulary
[3] zh = Chinese sentence
[4] rt = reference translation

### 5.1.2. Data Conversion

Upon completion of the instructional template, data acquisition will begin using CPTEC. The training data comprises two primary attributes ("*prompt*" and "*message*") presented in JSON format.

The attribute "*message*" comprises an array containing two distinct data: input and output. The attribute "*role*" should be configured to "*user*" to denote input data and "*assistant*" to designate output data.

The following JSON file exemplifies the conversion of a single CPTEC record into three distinct training data records. The attribute "*prompt*" represents the instructional directive, while the attribute "*messages*" contains the input data (Chinese sentences) alongside their corresponding reference translations. The correspondence between these data fields is detailed in **Table 7**.

```
{
  "prompt": "...into elementary Portuguese..." +
            "...(alternative vocabularies) ..." +
            "...A1−A2...",
  "messages": [
    {"role": "user", "content": "Chinese sentence"},
    {"role": "assistant", "content":
     "elementary level reference translation"}]
}, {
  "prompt": "...into intermediate Portuguese..." +
            "...(alternative vocabularies) ..." +
            "...B1−B2...",
  "messages": [
    {"role": "user", "content": "Chinese sentence"},
    {"role": "assistant", "content":
     "intermediate level reference translation"}]
}, {
  "prompt": "...into advanced Portuguese..." +
            "...(alternative vocabularies) ..." +
            "...C1−C2...",
  "messages": [
    {"role": "user", "content": "Chinese sentence"},
    {"role": "assistant", "content":
     "advanced level reference translation"}]
}, {
...
}
```

## 5.2. Fine-Tuning the LLM

Upon completion of data preparation, the subsequent step involves fine-tuning the LLM. Despite the fact that sequence variations may exist between models, the overarching process remains relatively consistent.

**Algorithm 2** outlines the logical flow of the LLM fine-tuning process. Initially, a model and tokenizer are instantiated. Following this, the source data is split into training and testing sets, and a variable, designated "*args*", is established to encapsulate the training parameters. Subsequently, the fine-tuning process is initiated utilizing the model, tokenizer, dataset, and specified arguments. Finally, the trained model is saved to the local file system for reuse.

**Algorithm 2** Programming Logic of fine-tuning a Large Language Model

**procedure** FT_LLM
  $model_{pt} \leftarrow LLM_{pt}$
  $tokenizer \leftarrow Pretrained(model_{pt})$
  $data \leftarrow LoadData("json",$
    $D_{CPTEC}.split(train \leftarrow 0.9,$
    $test \leftarrow 0.1))$ ▷ D: CPTEC Dataset
  $model_{llm} \leftarrow Pretrained(model_{pt})$
  $args \leftarrow TrainArg(epoch, batch, lrate)$
  $trainer \leftarrow Trainer(model_{llm},$
    $tokenizer, data, args)$
  $trainer.train()$ ▷ Fine-tuning
  $trainer.savemodel(localpath)$

## 5.3. Result Analysis

The process of fine-tuning an LLM is relatively rapid, facilitated by the computational power of modern *Graphics Processing Units* (GPUs). The duration of the process depends largely on the volume of training data. Moreover, since LLMs are usually tuned rather than built from scratch, a limited number of training epochs is generally sufficient. Iterative modifications to the instructions can be employed to optimize the output and achieve the desired results.

To evaluate the performance of the fine-tuned model, we tasked it with translating Chinese sentences from the CPTEC dataset into various levels of Portuguese. Subsequently, a comparison between LLMs without fine-tuning (*wo*) and those with fine-tuning (*ft*) yielded notable findings. For ease of analysis, the original Chinese text (*ot*), translated into English, is also provided.

### 5.3.1. BLEU Value

A more straightforward approach to evaluating results prior to fine-tuning involves comparing them using the BLEU value. Since the training process uses the CPTEC data, it is expected that the fine-tuned model will get higher BLEU values. We can calculate the BLEU value using the BLEU score module of the *Natural Language Toolkit* (NLTK) library in Python. BLEU values were calculated using 20 sentences for each of the 12 categories.

The results indicate a BLEU value of 0.45 in average prior to fine-tuning, which increases to 0.63 in average after the fine-tuning process. Some categories (arts, literacies, and politics) cannot be improved as much as others. It may be that the current LLM is unable to cover diverse content in a particular category.

However, accurately quantifying the degree of translational optimization achieved by this 0.18-point improvement remains a challenge. Therefore, we conduct an in-depth analysis of the text together with our bilingual teachers.

### 5.3.2. Usage of Lexicon

The fine-tuned model translates sentences using a richer vocabulary, including words that come from alternative vocabularies and using CEFR-level vocabulary. However, the LLM without fine-tuning uses the suggested words directly and cannot convey the same depth or nuance.

For example, phrases that have not been fine-tuned often use simpler words. One translation states, "*...as tradições históricas de cada sítio...*" which uses "*...cada sítio...*" instead of a more expressive term. In contrast, the fine-tuned model improves this by saying, "*...Ela também inclui tradições históricas profundas e características geográficas...*" using "*...**características geográficas**...*" to add depth.

ot: China's culinary culture has a long history, which is not only reflected in the richness of ingredients and a variety of cooking techniques, but also contains a deep historical tradition and regional characteristics.

*wo: A cultura da comida chinesa é muito longa, mostra os ricos alimentos, as suas formas de cozinhar e as tradições históricas de **cada sítio**.*

*ft: A cultura alimentar chinesa é muito rica e vai além de*

*apenas materiais alimentares ricos e técnicas de cozinhar. Ela também inclui tradições históricas profundas e* ***características geográficas****.*

As another example, a comparison of the fine-tuned sentences with the untuned ones reveals a clear improvement in word choice. In the following example, the untuned model uses "*...preocupação das pessoas...*" which sounds negative, while the fine-tuned version says "*...****consciência sobre a importância****...*" suggesting a better and positive understanding, that aligns better with the tone of the original sentence. Finally, the verb group "*...****surgiu e se consolidou****...*" in the fine-tuned version shows stronger acceptance of the idea, making the sentence clearer and more engaging.

ot: With an increasing focus on health and nutrition, the concept of functional foods and superfoods has emerged as an important part of modern food culture.

*wo: Com o aumento da* ***preocupação*** *das pessoas com a saúde e a nutrição, o conceito de alimentos funcionais e superalimentos* ***surgiu como*** *uma parte importante da cultura alimentar moderna.*

*ft: Com o crescimento da* ***consciência sobre a importância*** *da saúde e nutrição, o conceito de alimentos funcionais e super-alimentos* ***surgiu e se consolidou como*** *um componente crucial da gastronomia moderna.*

### 5.3.3. Classification of Sentence Difficulty

Without fine-tuning, the vocabulary employed in the following sentences remains at a comparable level of complexity. This consistency makes it difficult to differentiate between elementary and advanced linguistic difficulty.

ot: Technologists are looking at how online learning platforms and virtual classrooms can be utilized to provide flexible and diverse learning opportunities for students of all backgrounds.

A: *Os especialistas em tecnologia estão a estudar como usar a plataforma de aprendizagem online e salas de aula virtuais para oferecer oportunidades de aprendizagem flexíveis e variadas para estudantes de diferentes contextos.*

B: *Os especialistas em tecnologia estão a estudar como utilizar a plataforma de aprendizagem on-line e salas de aula virtuais para proporcionar oportunidades de aprendizagem flexíveis e variadas para estudantes de diferentes origens.*

C: *Os especialistas em tecnologia estão a desenvolver estratégias para utilizar a plataforma de aprendizagem on-line e salas de aula virtuais, permitindo que os estudantes de diferentes contextos tenham acesso a oportunidades de aprendizagem flexíveis e variadas.*

In the following fine-tuned version, the sentences exemplify varying levels of linguistic complexity and sophistication, illustrating the progression in vocabulary and structure. In sentence A, the words are simple and the sentence is easy to follow. This clarity makes it easy for more people to understand without feeling confused. In sentence B, the choice of the verb "*investigar*" (to investigate) instead of "*estudar*" (to study) introduces a layer of depth and precision. This level employs more nuanced language while remaining comprehensible, allowing for a richer expression of the ideas presented. Including the phrase "*...****uma variedade de oportunidades de aprendizagem flexíveis****...*" (a variety of flexible learning opportunities) further enhances the sophistication of the sentence. Sentence C represents the highest level of complexity, featuring complex syntax and advanced terminology. Words like "*desenvolver*" (to develop) and "*estratégias inovadoras*" (innovative strategies) significantly increase formality expression. The more elaborate structure, including clauses that provide context ("*...****permitindo que alunos de diferentes origens acedam a oportunidades de aprendizagem flexíveis e diversificadas***."), demonstrates a sophisticated command of the language, catering to a more discerning audience.

A: *Os especialistas em tecnologia estão a estudar como usar a plataforma de aprendizagem online e as salas de aula virtuais para oferecer diferentes oportunidades de aprendizagem para estudantes de várias origens.*

B: *Os especialistas em tecnologia estão a* ***investigar*** *como utilizar a plataforma de aprendizagem online e salas de aula virtuais para proporcionar* ***uma variedade de oportunidades de aprendizagem flexíveis*** *para alunos de diversos contextos.*

C: *Os especialistas em tecnologia estão a* ***desenvolver estratégias inovadoras*** *para tirar partido da plataforma de aprendizagem online e das salas de aula virtuais,* ***permitindo que alunos de diferentes origens acedam a oportunidades de aprendizagem flexíveis e diversificadas****.*

### 5.3.4. Miscellaneous

The sentences without fine-tuning are longer and have more complex phrases, which can make it harder to read. For instance, it says, "*...considerada benéfica para a longevidade...*" (considered beneficial for longevity) and then adds a lot of details, like "*...fornece uma boa fonte de macronutrientes essenciais, como proteínas, carboidratos e gorduras saudáveis...*", which stretches the sentence.

In contrast, the fine-tuned version is more concise. Using a similar length, this version presents two statements. It uses "***além disso***" (furthermore) to advance the idea without overloading the first sentence. This transition helps maintain clarity and flow while effectively conveying the message.

ot: The Mediterranean diet, rich in olive oil, fresh fruits and vegetables, fish, and whole grains, is recognized as contributing to a long and healthy life and is revered by healthy eating enthusiasts around the world.

*wo: A dieta mediterrânea, rica em azeite, vegetais frescos, peixes e grãos inteiros, é* ***considerada benéfica para a longevidade*** *e é apreciada por fãs de alimentação saudável em todo o mundo,* ***pois fornece*** *uma boa fonte de macronutrientes essenciais,* ***como*** *proteínas, carboidratos e gorduras saudáveis, que contribuem para um metabolismo equilibrado e uma boa saúde.*

*ft: A dieta mediterrânea, rica em azeite, vegetais frescos, peixes e grãos inteiros, é considerada uma opção útil para uma vida longa e saudável, pois fornece uma alimentação equilibrada que beneficia o metabolismo e a longevidade.* ***Além disso****, é apreciada por fãs de alimentos saudáveis em todo o mundo, que buscam uma dieta rica em frutas, proteínas e macronutrientes essenciais para uma boa saúde.*

An untuned LLM will generate additional information or repeat the translation when it encounters a sentence it does not understand. They often appear in most LLMs and are technically called "hallucination". For example, in the untuned version, the sentence is long and includes unnecessary details, making it harder to follow, and it says, "*...um ambiente favorável para o comércio internacional...*" (a favorable environment for international trade) and mentions "*a competitividade das empresas*" (the competitiveness of companies), which is not part of the original Chinese sentence. These added phrases make the translation more complicated and less clear. In the fine-tuned version, the message is clear and straight to the point. It says that "*The liberalization of trade and globalization have provided more opportunities and markets for international trade and promoted economic development and cooperation*". This version focuses on the main ideas without adding unnecessary details.

ot: Trade liberalization and globalization have provided more opportunities and markets for international trade and promoted economic development and cooperation.

*wo: A liberalização do comércio e a globalização proporcionam um ambiente* ***favorável*** *para o comércio internacional, desencadeando um processo de desenvolvimento e cooperação econômica que, por sua vez, contribuem para a prosperidade e* ***a competitividade das empresas. Além disso, essas tendências permitem a diversificação das economias e o intercâmbio cultural, tecnológico e comercial, o que é fundamental para o crescimento inovador e sustentável****.*

*ft: A liberalização do comércio e a globalização proporcionaram mais oportunidades e mercados para o comércio internacional e promoveram o desenvolvimento económico e a cooperação.*

An LLM that is not fine-tuned may lead to translation errors due to a misinterpretation of the meaning. In the original Chinese sentence, the focus is on how the boy's life changes due to the book. This translation introduces unnecessary details like finding the book in an antique shop and changing his course, "*...em uma pequena loja de antiguidades...*", which are not in the original and misinterpret the story's emphasis. In contrast, the fine-tuned version accurately reflects the original meaning that the boy's fate changes due to the mysterious book.

ot: The story begins in a small village where the fate of an ordinary boy is changed by a mysterious book.

*wo: O rapaz da localidade encontrou um livro antigo* ***em uma pequena loja de antiguidades*** *e, devido a sua curiosidade,* ***decidiu alterar seu rumo e ler o conteúdo****.*

*ft: A história começa em uma pequena aldeia, onde o destino de um menino comum começa a mudar após a descoberta de* ***um livro misterioso****.*

# 6. Conclusion

Recent advances in AI and LLM have significantly improved the quality of machine translation. However, the

output of machine translation often requires manual review and modification before formal deployment. This article explores the real-world application of our AI-based language learning application that automatically assesses translation and interpretation techniques. This computer-assisted assessment method shows potential for reinforcing language acquisition. To effectively bolster the skills of Chinese-Portuguese translators, our team successfully developed CPTEC and a hybrid grade system. Therefore, mobile software developers can integrate CPTEC and a hybrid grade system to positively influence learning motivation and anxiety.

We demonstrate the process of fine-tuning the LLM using the CPTEC to validate its ability to enhance translation quality across multiple dimensions, including accuracy, fidelity, fluency, readability, acceptability, and usability. Therefore, we make this practice dataset available to open source communities as a public contribution.

Although this article demonstrates a lot of research work in terms of Chinese-Portuguese translation quality, automatic evaluation methods, practical dataset construction, and LLM validation, there is still room for further innovation. Future research should explore the integration of tasks into mobile software designed to encourage self-regulation, collaborative learning, and metacognitive abilities. Such complementary developments would enhance the functionality of mobile software by providing more personalized feedback, thereby supporting users' progress and practice. In the same token, the creation of a diverse language corpus would benefit a wider range of translators. Furthermore, the implementation of voice and speech-to-text modules could facilitate access for users using mobile devices without the need for manual input.

Therefore, we will continue to analyze feedback data derived from our mobile app and visualize user activity patterns to identify significant insights. Furthermore, we will integrate CPTEC with advanced AI and LLM technologies to develop novel software solutions specifically designed to benefit our Chinese-Portuguese translators.

## Author Contributions

Conceptualization, L.M.H. and Y.S.; methodology, L.M.H.; software, L.M.H.; validation, L.M.H., Y.S. and M.L.; formal analysis, L.M.H.; investigation, Y.S. and M.L.; resources, Y.S. and M.L.; data curation, Y.S. and M.L.; writing—original draft preparation, L.M.H. and Y.S.; writing—review and editing, S.K.I.; supervision, S.K.I.; project administration, S.K.I.; funding acquisition, Y.S.. All authors have read and agreed to the published version of the manuscript.

## Funding

This work was supported by Macao Polytechnic University grant number [RP/FCA-07/2023].

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

In this work, we created a self-study material (CPTEC) and released this material to the AI community (https://huggingface.co/datasets/edmond5995/CPTransExercise).

## Acknowledgments

Data has always been regarding as a valuable asset in both academia and business. For this, we would like to thank the Corpus Team of *Macao Polytechnic University* (MPU) for their assistance in the production of CPTEC mentioned in Section 3.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

[1] Sun, Y., Hoi, L.M., Im, S.K., 2023. Constructing the evaluation index system of Chinese-portuguese machine translation using the delphi and analytic hierarchy process methods. Proceeding of the 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA); 27–29 January

2023; Shenyang, China. pp. 190–195. DOI: https://doi.org/10.1109/ICPECA56706.2023.10075920

[2] Deci, E.L., Ryan, R.M., 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry. 11(4), 227–268. DOI: https://doi.org/10.1207/S15327965PLI1104_01

[3] Kukulska-Hulme, A., 2020. Mobile-assisted language learning. In: Chapelle, C.A., Aijmer, K., Angelelli, C.V., et al. (eds.). The Encyclopedia of Applied Linguistics. John Wiley & Sons, Ltd: Hoboken, NJ, USA. pp. 1–9. DOI: https://doi.org/10.1002/9781405198431.wbeal0768.pub2

[4] Hoi, L.M., Sun, Y., Ke, W., et al., 2023. Visualizing the behavior of learning european portuguese in different regions of the world through a mobile application. IEEE Access. 11, 113913–113930. DOI: https://doi.org/10.1109/ACCESS.2023.3324390

[5] HKYWCA, 2025. Study finds strong parental influence on children's electronic device usage. Available from: https://www.ywca.org.hk/Press/Children-Electronic-Device-Usage (cited 23 January 2025).

[6] Gentile, D.A., Reimer, A.R., Nathanson, A.I., et al., 2014. Protective effects of parental monitoring of children's media use a prospective study. JAMA Pediatrics. 168(5), 479–484. DOI: https://doi.org/10.1001/jamapediatrics.2014.146

[7] Qayyum, A., Shahid, R., Fatima, M., 2024. The effect of excessive smartphone use on child cognitive development and academic achievement: A mixed method analysis. Annals of Human and Social Sciences. 5(3), 166–181. DOI: https://doi.org/10.35484/ahss.2024(5III)16

[8] Yang, J., Por, L.Y., Leong, M.C., et al., 2023. The potential of chatgpt in assisting children with down syndrome. Annals of Biomedical Engineering. 51(12), 2638–2640. DOI: https://doi.org/10.1007/s10439-023-03281-3

[9] CATTI, 2025. CATTI Integrated Service Platform. Available from: https://www.catticenter.com/cattiptyyksdg (cited 30 June 2025).

[10] Bittner, H., 2019. Evaluating the Evaluator: A Novel Perspective on Translation Quality Assessment, 1st ed. Routledge: Oxfordshire, UK.

[11] Peña Pollastri, A.P., 2008. Evaluation criteria for the improvement of translation quality. In: Forstner, M., Schmitt, P.A. (eds.). CIUTI-Forum 2008: Enhancing Translation Quality: Ways, Means, Methods. Peter Lang: Bern, Switzerland. pp. 239–260.

[12] Martínez Mateo, R., 2014. A deeper look into metrics for translation quality assessment (tqa): A case study. Language and Linguistics. 49, 73–93. DOI: https://doi.org/10.26754/ojs_misc/mj.20148792

[13] Jiang, L., Sun, Y., Hoi, L.M., 2024. Defining the indicator system for awe for portuguese classes in chinese higher education. Proceedings of the 2024 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC); 18–20 October 2024; Guangzhou, China. pp. 206–211. DOI: https://doi.org/10.1109/CyberC62439.2024.00043

[14] Machine Translate Foundation, 2024. Ninth conference on machine translation. Available from: https://machinetranslate.org/wmt24 (cited 26 November 2024).

[15] Papineni, K., Roukos, S., Ward, T., et al., 2002. Bleu: a method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics; 7–12 July 2002; Philadelphia, PA, USA. pp. 311–318. DOI: https://doi.org/10.3115/1073083.1073135

[16] Lin, C.Y., 2004. Rouge: A package for automatic evaluation of summaries. Proceedings of the ACL-04 Workshop; 25–26 July 2004; Barcelona, Spain. pp. 74–81.

[17] Radford, A., Narasimhan, K., Salimans, T., et al., 2018. Improving language understanding by generative pre-training. OpenAI. Available from: https://openai.com/index/language-unsupervised (cited 29June 2018).

[18] Radford, A., Wu, J., Child, R., et al., 2019. Language models are unsupervised multitask learners. OpenAI blog. 1(8), 9.

[19] Brown, T.B., Mann, B., Ryder, N., et al., 2020. Language models are fewshot learners. Proceedings of the 34th International Conference on Neural Information Processing Systems; 6–12 December 2020; Vancouver, Canada. pp. 1877–1901.

[20] OpenAI, 2021. Gpt-3 powers the next generation of apps. Available from: https://openai.com/index/gpt-3-apps (cited 27 March 2021).

[21] OpenAI, 2021. Introducing chatgpt. Available from: https://openai.com/index/chatgpt (cited 30 November 2021).

[22] Achiam, J., Adler, S., Agarwal, S., et al., 2023. Gpt-4 technical report. arXiv. DOI: https://doi.org/10.48550/arXiv.2303.08774

[23] Wu, T., He, S., Liu, J., et al., 2023. A brief overview of chatgpt: The history, status quo and potential future development. IEEE/CAA Journal of Automatica Sinica. 10(5), 1122–1136. DOI: https://doi.org/10.1109/JAS.2023.123618

[24] Mughal, N., Mujtaba, G., Shaikh, S., et al., 2024. Comparative analysis of deep natural networks and large language models for aspect-based sentiment analysis. IEEE Access. 12, 60943–60959. DOI: https://doi.org/10.1109/ACCESS.2024.3386969

[25] Moradi Dakhel, A., Nikanjam, A., Khomh, F., et al., 2024. Generative AI for Effective Software Development. Springer Nature: Cham, Switzerland.

[26] Alammar, J., Grootendorst, M., 2024. Hands-On Large Language Models: Language Understanding and Generation, 1st ed. O'Reilly Media: Sebastopol, CA, USA.

[27] Tunstall, L., Werra, L.v., Wolf, T., 2022. Natural Language Processing with Transformers: Building Lan-

guage Applications with Hugging Face, 1st ed. O'Reilly Media: New York, NY, USA.

[28] Hu, E.J., Shen, Y., Wallis, P., et al., 2022. Lora: Low-rank adaptation of large language models. Proceedings of the Tenth International Conference on Learning Representations (ICLR 2022); 25–29 April 2022; Virtual. pp. 1–13.

[29] Dettmers, T., Pagnoni, A., Holtzman, A., et al., 2023. Qlora: efficient finetuning of quantized llms. Proceedings of the 37th International Conference on Neural Information Processing Systems; 10–16 December 2023; New Orleans, LA, USA. pp. 10088–10115.

[30] Council of Europe, 2020. Common European Framework of Reference for Languages: Learning, Teaching, assessment: Companion volume, 1st ed. Cambridge University Press: London, UK.

[31] Nida, E.A., 1964. Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating. E.J. Brill: Leiden, Netherlands.

[32] Pym, A., 2023. Exploring Translation Theories, 3rd ed. Routledge: London, UK.

[33] Bassnett, S., Lefevere, A., 1990. Translation, History and Culture, 1st ed. Cassell: London, UK.

[34] Chinese-Portuguese-English Machine Translation Laboratory (CPELab), 2024. Chinese-Portuguese Translation Exercise Corpus (CPTEC). Available from: https://huggingface.co/datasets/edmond5995/CPTransExercise (cited 15 December 2024).

[35] Vohra, D., 2016. Apache parquet. In: Vohra, D. (ed.). Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools. Apress: Berkeley, CA, USA. pp. 325–335.

[36] Oxford Dictionaries, 2015. Oxford Portuguese Dictionary, 1st ed. Oxford University Press: Oxford, UK.

[37] Wiktionary, 2023. Portuguese lemmas. Available from: https://en.wiktionary.org/wiki/Category:Portuguese_lemmas (cited 12 April 2023).

[38] Lommel, A., Uszkoreit, H., Burchardt, A., 2014. Multidimensional quality metrics (mqm): A framework for declaring and describing translation quality metrics [in Catalan]. Revista Tradumàtica: Tecnologies de la Traducció. 12, 455–463. DOI: https://doi.org/10.5565/rev/tradumatica.77

[39] Bojar, O., Chatterjee, R., Federmann, C., et al., 2016. Findings of the 2016 conference on machine translation. Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers; 11–12 August 2016; Berlin, Germany. pp. 131–198. DOI: https://doi.org/10.18653/v1/W16-2301

[40] Toral, A., Sánchez-Cartagena, V.M., 2017. A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1; 3–7 April 2017; Valencia, Spain. pp. 1063–1073.

[41] Hassan, H.M., Galal-Edeen, G.H., 2017. From usability to user experience. Proceedings of the 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS); 24–26 November 2017; Okinawa, Japan. pp. 216–222. DOI: https://doi.org/10.1109/ICIIBMS.2017.8279761

[42] Kri, R., Sambyo, K., 2024. Comparative study of low resource digaru language using smt and nmt. International Journal of Information Technology. 16(4), 2015–2024. DOI: https://doi.org/10.1007/s41870-024-01769-2

[43] Sälevä, J., Lignos, C., 2021. The effectiveness of morphology-aware segmentation in lowresource neural machine translation. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop; 19–23 April 2021; Virtual. pp. 164–174. DOI: https://doi.org/10.18653/v1/2021.eacl-srw.22

[44] Zhang, M., Li, Z., Fu, G., et al., 2019. Syntax-enhanced neural machine translation with syntax-aware word representations. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2–7 June 2019; Minneapolis, MN, USA. pp. 1151–1161. DOI: https://doi.org/10.18653/v1/N19-1118

[45] McDonald, S.V., 2022. Accuracy, readability, and acceptability in translation. Applied Translation. 16(2), 1–9. DOI: https://doi.org/10.51708/apptrans.v14n2.1238

[46] Callison-Burch, C., Osborne, M., Koehn, P., 2006. Re-evaluation the role of bleu in machine translation research. Proceedings of the 11th Conference of the European Chapter of the Association for Compuational Linguistics; 3–7 April 2006; Trento, Italy. pp. 249–256.

[47] Su, J., Chen, J., Jiang, H., et al., 2021. Multi-modal neural machine translation with deep semantic interactions. Information Sciences. 554, 47–60. DOI: https://doi.org/10.1016/j.ins.2020.11.024

[48] Muftah, M., 2022. Machine vs human translation: a new reality or a threat to professional arabic—english translators. PSU Research Review. 8(2), 484–497. DOI: https://doi.org/10.1108/PRR-02-2022-0024

[49] Chesterman, A., 2016. Memes of Translation: The Spread of Ideas in Translation Theory. John Benjamins: Amsterdam, The Netherlands.

[50] Python Software Foundation, 2025. locale internationalization services. Available from: https://docs.python.org/3/library/locale.html (cited 25 May 2025).

[51] Hoi, L.M., Ke, W., Im, S.K., 2023. Corpus database management design for chinese-portuguese bidirectional parallel corpora. Proceedings of the 2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI); 26–28 May 2023; Taiyuan, China. pp. 103–108. DOI: https:

//doi.org/10.1109/CCAI57533.2023.10201319

[52] de Azeredo, J.C., 2012. Houaiss Verb Conjugation Dictionary, 1st ed [in Portuguese]. Publifolha: São Paulo, Brazil.