FACULTAD DE TRADUCCIÓN Y DE INTERPRETACIÓN
GRADO DE TRADUCCIÓN E INTERPRETACIÓN

TRABAJO DE FINAL DE GRADO
Curso 2025-2026

Localising the MMLU: Multilingual Style Guide Development and Translation Problem Analysis for Preserving Cross-Linguistic Benchmark Integrity

Sabela Calvet Álvarez
1599775

TUTORA
PILAR SÁNCHEZ-GIJÓN

Barcelona, 31 de mayo de 2026

UAB
Universitat Autònoma de Barcelona

Tal vez todo sea una inmensa ilusión. Quién sabe. Pero no lograremos la verdad que esconde dándole la espalda, sino hundiéndonos en ella.

ANTONIO BUERO VALLEJO, *La Fundación*

Credits Page

Title:

Localising the MMLU: Multilingual Style Guide Development and Translation Problem Analysis for Preserving Cross-Linguistic Benchmark Integrity

Localización del MMLU: Desarrollo de una guía de estilo multilingüe y análisis de problemas de traducción para preservar la integridad interlingüística del *benchmark*

Localització del MMLU: Desenvolupament d'una guia d'estil multilingüe i anàlisi de problemes de traducció per a preservar la integritat interlingüística del *benchmark*

Author: Sabela Calvet Álvarez

Tutor: Pilar Sánchez-Gijón

University: Universitat Autònoma de Barcelona (UAB)

Studies: Translation and Interpreting

Academic year: 2025/2026

Key Words

Benchmark dataset, Directorate-General for Translation of the European Commission, European Master's in Translation, LLM, localisation, MMLU dataset, multilingual style guide

Benchmark dataset, Dirección General de Traducción de la Comisión Europea, Máster Europeo en Traducción, LLM, localización, MMLU dataset, guía de estilo multilingüe

Benchmark dataset, Direcció General de Traducció de la Comissió Europea, Màster Europeu en Traducció, LLM, localització, MMLU dataset, guia d'estil multilingüe

Abstract

The MMLU benchmark emerged as a method for evaluating the general knowledge of Large Language Models (LLMs), driven by their unprecedented advancement in recent years. Today, the European Union (EU) aspires to develop a benchmark dataset building on the MMLU but localised to reflect European values. To this end, the present study develops a multilingual style guide to support the development of the project, alongside a subsequent translation problem analysis aimed at exploring aspects not covered by the

style guide and proposing a course of action for the development of future style guides in projects of this nature. The style guide was developed drawing on the EU Interinstitutional Style Guides, while the translation problem analysis was carried out on the main problem types that emerged during the translation process. Based on the style guide, the problems encountered, and the solutions adopted, it is established that while the guide provided a solid foundation for the linguistic aspects during the translation process, it proved insufficient to address the full range of problems posed by the specific nature of the MMLU as a text type but will provide a valuable basis for developing future versions.

El *benchmark dataset* MMLU surgió como método de evaluación de conocimiento general de *Large Language Models* (LLMs) motivado por su avance sin precedentes. Hoy, la Unión Europea (UE) aspira a desarrollar un *benchmark dataset* a raíz del MMLU, pero localizado para reflejar los valores europeos. Para ello, el presente trabajo desarrolla una guía de estilo multilingüe para apoyar el desarrollo de dicha traducción, junto con un análisis de problemas de traducción posterior, destinado a explorar aquellos aspectos que no han quedado reflejados en la guía de estilo y a proponer un curso de acción para el desarrollo de futuras guías en proyectos de esta naturaleza. La guía de estilo se lleva a cabo tomando como base las Guías de Estilo Interinstitucionales de la UE, mientras que el análisis de problemas se realiza sobre los principales tipos de problemas que surgen durante el proceso de traducción. A partir de a la guía de estilo, los problemas encontrados y las soluciones adoptadas, se establece que, si bien la guía proporcionó una base sólida para los aspectos lingüísticos durante el proceso de traducción, resultó insuficiente para abordar la totalidad de los problemas planteados por la naturaleza específica del MMLU como tipo textual, aunque servirá como base valiosa para el desarrollo de futuras versiones.

El *benchmark dataset* MMLU va sorgir com a mètode d'avaluació del coneixement general dels *Large Language Models* (LLMs), impulsat pel seu avanç sense precedents en els darrers anys. Avui, la Unió Europea (UE) aspira a desenvolupar un *benchmark dataset* basat en el MMLU, però localitzat per reflectir els valors europeus. Amb aquesta finalitat, el present treball elabora una guia d'estil multilingüe per donar suport a aquesta traducció, juntament amb una anàlisi de problemes de traducció posterior, destinada a explorar els aspectes no recollits en la guia i a proposar un curs d'acció per a futures guies en projectes d'aquesta naturalesa. La guia d'estil es basa en les Guies d'Estil Interinstitucionals de la UE, mentre que l'anàlisi de problemes s'efectua sobre els principals tipus de problemes sorgits durant el procés de traducció. A partir de la guia, els problemes trobats i les solucions adoptades, s'estableix que, si bé la guia va proporcionar una base sòlida per als aspectes lingüístics durant el procés de traducció, va resultar

insuficient per abordar la totalitat dels problemes plantejats per la naturalesa específica del MMLU com a tipus textual, tot i que constituirà una base valuosa per a l'elaboració de futures versions.

TABLE OF CONTENTS

1 INTRODUCTION ..... - 9 -

1.1. Motivation and Justification ..... - 9 -

1.2. Objectives and Research Questions ..... - 9 -

1.3. Outline of the Present Study ..... - 10 -

2. LITERATURE REVIEW ..... - 11 -

2.1. What are LLMs ..... - 11 -

2.1.1. Key Dimensions of LLM Evaluation ..... - 12 -

2.1.2. Key LLM Evaluation Criteria ..... - 12 -

2.1.3. Key Methodologies Used to Test LLMs ..... - 13 -

2.2. The MMLU Dataset ..... - 15 -

2.3. What Text Type does the MMLU belong to? ..... - 15 -

2.4. Translation Strategy and Techniques ..... - 17 -

2.5. Localisation and Internationalisation ..... - 18 -

2.6. Style Guides ..... - 19 -

2.6.1. Purpose ..... - 19 -

2.6.2. How to Build a Style Guide ..... - 19 -

2.6.3. About the EU Guidelines and the Interinstitutional Style Guide ..... - 20 -

3. METHODOLOGY ..... - 23 -

3.1. Project Overview ..... - 23 -

3.1.1. Description and Goal ..... - 23 -

3.1.2. Translation Brief ..... - 23 -

3.1.2.1. Previous Translation of the MMLU ..... - 23 -

3.1.2.2. Current Translation Brief ..... - 24 -

3.2. Objectives and Scope ..... - 25 -

3.3. Analysis of the Source Material ..... - 26 -

3.4. Identification of Relevant DGT Guidelines and Analysis ..... - 27 -

3.5. Design and Justification of the Project Style Guide ..... - 31 -

3.6. Translation Problems Beyond the Current Style Guide ..... - 33 -

4. DISCUSSION ..... - 44 -

4.1. Assessment of Translation Solutions ..... - 44 -
4.2. Recommendations for Future Editions ..... - 50 -
5. CONCLUSION ..... - 51 -
5.1. Revisiting Research Questions and Objectives ..... - 51 -
5.2. Future Lines of Research ..... - 54 -
6. BIBLIOGRAPHY ..... - 55 -
7. APPENDIX ..... - 60 -

LIST OF TABLES

Table 1. Annotated index of Parth Three of the Interinstitutional Style Guide. .....- 29 -
Table 2. Selection of entries from Part Three of the Interinstitutional Style Guide. ..- 29 -
Table 3. Selected chapters and entries from Part Four of the Spanish version of the Interinstitutional Style Guide.....- 31 -
Table 4. First row of project's style guide.....- 32 -
Table 5. Problem analysis template. .....- 35 -
Table 6. Translation problem analysis and justification of author’s solutions. .....- 49 -

1 INTRODUCTION

In the last decade, the rapid evolution of Large Language Models (LLMs), driven by transformer models and their ability to pretrain on enormous amounts of text corpora, has brought about the need for a benchmark dataset that could properly evaluate the current LLMs' capabilities (Hendrycks, 2021). That is what led Hendrycks et al. in 2020 to develop the MMLU benchmark (Massive Multitask Language Understanding), a new and more challenging test for the state-of-the-art AI systems (Kumar, 2025). Today, the MMLU has become a reference benchmark.

The present study is framed within a project promoted by the AI Unit of the Directorate-General for Translation (DGT) of the European Commission in collaboration with the European Masters in Translation (EMT) network that consists of translating the MMLU into the 24 official European Union (EU) languages. In this is the first edition of the project the benchmark will be translated into 11 out of the 24 official EU languages. To do so, a multilingual group of master students among the different EMT member universities has been established.

1.1. Motivation and Justification

The motivation for this study arises from my direct involvement in the DGT-EMT MMLU Translation Project. Throughout my degree, I came to realise that translation is a discipline applicable across a wide range of fields. However, I would have never imagined how relevant it could be in the world of technology. That moved me to try and find a topic related to translation technologies and AI for my dissertation. That is when my tutor offered me to take part in this project. And, without knowing it, I was just about to delve into a topic that is not only fascinating, but one that I have been into almost since I started this degree. Furthermore, the aim of the project was unique and the possibility to collaborate in a European project was one that I could not resist.

As of why the project came to be, the DGT-EMT initiative arises from the need to create an LLM evaluation system of European nature (for the original data is based on American English) and is aligned with European values of multilingualism and multiculturalism. The existence of a localized benchmark for every official EU language allows for the evaluation of current LLMs in said languages, helping determine their strengths and weaknesses. This will then lead to further development, promoting models that are as capable in a language with fewer speakers as they are in a widely spoken one, reducing the digital divide (Folaron, 2025) and protecting Europe's linguistic diversity (European Union, n.d.d).

1.2. Objectives and Research Questions

This study aims to develop a multilingual style guide that serves in the translation of the MMLU. In order to achieve that, the following specific objectives are set:

1. Develop a theoretical understanding of what benchmarks are, what they measure, and how they do so in order to understand the textual function of the text.
2. Develop a style guide based on official EU linguistic standards.
3. Identify to which extent the style guide meets the needs of translators by developing a systematic approach to analyse key translation problems found in the source text (ST).
4. Propose a course of action for the creation of a final style guide in future editions of the project.

These objectives are explored through the following research questions:

1. Does the decontextualised nature of the MMLU generate different translation problems to those found in conventional translation projects?
2. Can a style guide as conventionally defined in translation studies be used to tackle assignments such as the translation of the MMLU?
3. To what extent is the MMLU translatable without compromising the difficulty level and evaluative purpose for which it was originally designed?
4. How does the nature of the MMLU redefine the role of the translator and the decision-making process?

1.3. Outline of the Present Study

The present study will respond to the following structure:

First, the essential theoretical concepts to understanding the development of the project will be explored in the literature review. Here the technical concepts like LLM evaluation criteria and methodologies, benchmark dataset and MMLU will be explained. Other fundamental concepts like style guide, localisation or translation techniques will also be tackled, for they are an essential part for understanding the next section.

Second, the methodology will explore in detail how the style guide was created and the reasoning behind it, as well as a brief overview of the project. Then, the translation problem analysis will dive into the systematic methodology followed to describe each problem, followed by the analysed problems themselves.

Lastly, the discussion covers a further reasoning on why certain solutions were adopted and explores recommendations for future editions, while the conclusion gives an answer to whether the objectives were met and tries to answer the research questions posed above.

2. LITERATURE REVIEW

2.1. What are LLMs

According to Stryker (2025):

> Large language models (LLMs) are a category of deep learning models trained on immense amounts of data, making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks. LLMs are built on a type of neural network architecture called a transformer which excels at handling sequences of words and capturing patterns in text.

LLMs encompass a wide range of capabilities, including programming and technical skills, tool and agent use, and multimodality (combining language with formats such as images, audio, and video). Of particular relevance to the present study, however, are three core capabilities: the ability to generate language, this is, produce textual content; the capacity to understand and transform texts through natural language processing (NLP); and the ability to carry out guided reasoning and task resolution, simulating reasoning processes by identifying patterns, inferring plausible continuations, and constructing meaningful outputs form training data and context input (Diego, 2026).

They emerged from years of incremental progress in NLP and machine learning, and they have played a central role in the rapid acceleration of AI innovation during the late 2010s and 2020s. Today, their use ranges from text generation, summarization and translation to code generation, sentiment analysis and AI assistance (Stryker, 2025).

Additionally, LLMs can be fine-tuned to specialise in a concrete task, domain or style. However, some larger trained (or, in this context, *pre-trained*, as data scientists prefer to call it to differentiate it from fine-tuning, for both are technically "training" (Bergmann, 2025)) models have reached such astonishing performance only with pre-training that it "removes the need for a large fine-tuning set" (Hendrycks et al., 2021).

To ensure the proper performance of LLMs, they are evaluated prior to their release. The importance of this process relies on the ability to mitigate underperformances in the output, which may manifest in any output modality but, for the purpose of this study, the focus is on linguistic outputs. Such underperformances include, but are not limited to:

Hallucinations: generating information that is incorrect, unfounded, or inconsistent with reality. The outputs may appear plausible, but they lack factual or logical grounding (Farnschläder, 2025).

Biases and Unsafety: inaccuracies or potentially harmful outputs derived from human prejudices embedded in a model's training data (Holdsworth, n.d.).

Inefficiency: the goal is to make the models faster and more efficient.

There are many evaluation standards that follow an established set of criteria, crucial for evaluating the quality of LLM outputs. They also aim to enhance model performance, guarantee accuracy, strengthen trust and reliability, mitigate ethical and

safety risks, and assess business worth (Joshi, 2025). Understanding these standards requires distinguishing between three interconnected elements: the key dimensions that define what is being evaluated, the criteria that provide the quantitative means to measure it, and the methodologies that determine how evaluation is conducted.

2.1.1. Key Dimensions of LLM Evaluation

Evaluating LLMs is a complex task, as their performance cannot be measured through a single criterion alone. Due to the wide range of tasks that these systems are expected to perform, current evaluation approaches tend to assess several core dimensions that reflect different aspects of model behaviour and capability. According to Zhuang et al. (2023), four core competencies can be identified when assessing LLMs:

- **Knowledge**. Understood as the information obtained by LLMs through vast training corpora in order to solve a wide range of tasks. It is typically divided into: linguistic knowledge, which includes grammatical, semantic, and pragmatic knowledge; and world knowledge, which refers to non-linguistic information that nonetheless help the audience interpret the meaning of what they are reading or listening (Zhuang et al., 2023).
- **Reasoning**: fundamental to the problem-solving process of LLMs. Despite that, there is no consensus on whether LLMs genuinely reason or merely stimulate it through statistical prediction. Regardless, reasoning performance can be objectively assessed through reasoning competencies (Zhuang et al., 2023).
- **Reliability**: measures the degree to which the outputs of an LLM can be trusted. It has become a central concern as LLMs are increasingly deployed in real-world applications. This dimension encompasses concepts such as hallucination, truthfulness, factuality, calibration and robustness, and overlaps closely with safety and generalization (Zhuang et al., 2023).
- **Safety**: it concerns the risks associated with LLM outputs when deployed in real-world applications. The statistical nature of these models introduces inherent and underlying risks. Evaluating this dimension remains challenging, as current evaluative frameworks are insufficient, and the perception of safety is largely subjective (Zhuang et al., 2023).

Measuring these four dimensions requires clear standards against which model outputs can be assessed. The following section introduces the key criteria used to determine what constitutes quality performance across each of these dimensions.

2.1.2. Key LLM Evaluation Criteria

Assessing the quality of LLM outputs requires a set of established criteria that reflect the key properties a model must demonstrate. The following represent the most widely

recognised standards in the literature for evaluating the general performance of LLMs, independently of any specific application or task:

- **Accuracy** refers to how well an LLM can produce desired results. High accuracy is crucial for generating dependable and impartial predictions, which in turn ensures that LLMs meet user requirements and follow instructions effectively (Hu & Zhou, 2024).
- **Bias and Fairness** include demographic parity (consistency in performance regardless of demography)**,** equal opportunity (ensuring models do not disproportionately fail any demographic groups) and counterfactual fairness (whether model's predictions remain stable when sensitive attributes are altered) (Karzhev, 2024).
- **Fluency and Coherence** evaluate grammatical quality and the text's logical structure and consistency, ensuring that the production is an easy to follow, clear, human-like output (Karzhev, 2024).
- **Relevance and Informativeness** evaluate if the output is relevant and provides practical information (Joshi, 2025).
- **Robustness**, refers to an LLM's ability to withstand mistakes, manipulations, or adversarial attacks, ensuring that its output remain dependable and consistent (Hu & Zhou, 2024). It comes hand in hand with safety.
- **Efficiency** refers to the capacity of a model to deliver only the information necessary for task completion, avoiding redundancy and minimising unnecessary cognitive load (Miller & Tang, 2025).

Some other evaluation criteria include safety, explainability and reasoning, and social norm. These are focused on the current main applications of LLMs: Writing Assistance, Information Retrieval, Commercial Use, and Personal Use (Liu et al., 2023).

These criteria establish a comprehensive framework for defining what constitutes quality in LLM outputs. However, identifying what needs to be evaluated is only the first step. The way these criteria are assessed in practice varies significantly depending on the approach adopted, which leads to the evaluation methodologies.

2.1.3. Key Methodologies Used to Test LLMs

Evaluating LLMs is a complex task for which no universally accepted practices have been established, and there remains a notable lack of consensus on which approaches are most appropriate. As a result, researchers and developers must select the methodology that best suits their specific application and goals. Despite this variability, three main methodological approaches to the assessment process can be identified in the literature: automatic evaluation, human evaluation, and model-based evaluation (commonly referred to as LLM-as-a-judge) (Abeysinghe & Circi, 2024).

Automatic evaluation relies on standardised metrics and tools to assess model performance without human intervention, making it a time-efficient and objective approach that minimises subjective bias (Chang et al., 2024).

Human evaluation involves human participation (such as experts, researchers, or end users) directly assessing the quality of the model outputs. While more resource-intensive, it provides more comprehensive feedback and is closer to real-world application scenarios (Chang et al., 2024).

As for LLMs-as-judges, they use LLMs themselves as evaluators, replacing both human judges and statistical metrics. This approach offers three key advantages over traditional methods: it adapts evaluation criteria to the specific task context rather than relying on fixed metrics; it generates interpretive, nuanced feedback that provides deeper insight into the model performance; and it offers a scalable and reproductible alternative to human evaluation, significantly reducing the time and cost involved (Li et al., 2024).

Automatic evaluation operates primarily through benchmark datasets as its core instrument. A benchmark dataset is a standardized set of tasks that enables the evaluation of LLM performance across specific capabilities, facilitating comparative analysis between models and establishing a baseline for measurement (Karzhev 2024; Hadi et al., 2025). The format in which tasks are presented varies across benchmarks and determines how model responses are assessed. Three main formats can be identified: multiple choice format, in which options are explicitly presented in the prompt and prefixed with labels such as A/B/C/D; cloze formulation, in which the model selects between options without these being explicitly provided in the prompt; and freeform generation, in which the model produces unrestricted response to a given prompt (Fourrier et al., 2025).

Benchmark datasets are the primary instrument of another commonly used methodological approach, one that shifts the focus from who conducts the evaluation to what is being assessed. According to Ni et al. (2025a), this methodology relies on benchmark classification based on their thematic scope, that is, the capabilities they are designed to assess. There are three main capabilities: linguistic core, which reflects the ability to measure generalised linguistic competence, going beyond surface level pattern matching to assess deeper aspects of language, including syntax, semantics, and pragmatics; knowledge, which reflects the ability to store and accurately retrieve large amounts of real-world information; and reasoning, which reflects the ability to reason across formal logic, commonsense inference, and applied problem-solving.

The MMLU, the object of this study which will be described in the following section, primarily classifies as a knowledge benchmark, though its design also incorporates a reasoning component, as several tasks require the application and inference of knowledge rather than mere factual recall.

2.2. The MMLU Dataset

The Massive Multitask Language Understanding (MMLU) benchmark dataset is one of the most widely used tools for evaluating LLMs. Designed to assess model performance as the percentage of correct answers across all tasks, it provides a standardised metric for comparing models, enabling the AI community to identify leading systems and analyse the strengths and limitations of each approach (Kumar, 2025).

The MMLU, according to Hendrycks et al. (2021), is built around four core design features, each serving a specific evaluative purpose. First, it comprises 15619 multiple-choice questions. This format allows for efficient and fully automated evaluation of a model's capabilities (Ni et al., 2025b). Second, these questions are distributed across 57 tasks spanning from science, technology, engineering, mathematics to humanities and social sciences. This thematic diversity goes beyond linguistic comprehension, providing a holistic view of model capabilities and facilitating the identification of blind spots that can guide future research (Kumar, 2025). Third, the MMLU adopts a zero- or few-shot evaluation setting, which distinguishes genuine learning form overfitted or format-driven responses, making the benchmark more challenging and closer to how knowledge is assessed in humans (Hendrycks et al., 2021). This is made possible by the emergence of pretrained LLMs, which no longer require fine-tuning to perform. Finally, its heterogeneous difficulty level, ranging from elementary to advanced questions, ensures that both general world knowledge and complex problem-solving ability are put to the test.

Among the risks associated with the MMLU, the first concerns the uncertainty origin of its questions. Hendrycks et al. (2021) say:

> The questions in the dataset were manually collected by graduate and undergraduate students from freely available sources online. These include practice questions for tests such as the Graduate Record Examination and the United States Medical Licensing Examination. It also includes questions designed for undergraduate courses and questions designed for readers of Oxford University Press books.

However, they are not fully transparent with the origin of the questions and texts, raising questions about the validity and reliability of its sources. This directly affects the process of determining whether the model actually reasons or just recalls information. The final risk concerns whether the benchmark truly adopts a universal approach or whether, on the contrary, it is embedded within a particular sociocultural context.

2.3. What Text Type does the MMLU belong to?

Given that the project in which this study is framed aims to translate the MMLU, it is worth considering what text type the benchmark belongs to. The content of the MMLU is, as previously explored, not a conventional text but a set of questions, each with four possible answers, and aimed at an LLM instead of a human audience. To comprehend the

text type to which it belongs it is essential to understand the bases to designing a benchmark. As Cao et al. (2025) outline, there is an initial designing phase in the creation of a benchmark, where the scope and the required capabilities are established. Only then can the actual construction of the benchmark begin, which includes data collecting, curation, and validation. An evaluation phase follows, so the benchmark is applied to a representative selection of LLMs, where consistency is achieved through settings (like prompts). Next, experimental results are analysed by comparing the models' performance, useful to establish correlations between models, improvement areas, and guide future improvements in LLM development. Finally, all materials are carefully prepared, and the benchmark is made open-accessible.

This issue can also be explored from a linguistic perspective. According to De Beaugrande and Dressler (1981), to be understood as a communicative event, a text must satisfy the seven standards of textuality, which reflect the different dimensions along which communication can succeed. Those standards are: cohesion and coherence, which cover the text itself, how the elements of the text, both superficial and subjacent, are connected among themselves; intentionality and acceptability, which cover the intentions with which the message is emitted and the expectations and receptiveness of the receptor; informativity, which covers the message itself; situationality, which covers the context (time- and space-wise) in which the communication act takes place; and intertextuality, which covers its relationship with other texts.

However, the MMLU does not check all those requirements, for it consists of a series of units that are neither cohesive not coherent between them; it does have a clear intentionality, just not the one De Beaugrande and Dressler initially thought of, but there is no way of measuring acceptability since it is not aimed at a human receptor; informativeness cannot be adequately measured since the questions are too brief; situationality is also unmeasurable in the circumstances described by De Beaugrande and Dressler, who did not contemplate the scenario of a "communicative situation" with a non-human receptor; and intertextuality, which is present. Overall, the authors did not create these standards with a situation like the one under analysis in mind.

Furthermore, the text does not respond to any other textual functions: it does not just convey factual information (not informative), it does not try and convince the receiver (not appellative), it does not try and reveal the emotions of the speaker (because there is no speaker, so it is not expressive), it does not focus on the aesthetic of the text (so it is not poetic), it does not try and clarify language (so it is not metalinguistic), and lastly, it does not focus on the communication channel (so it is not phatic). Its sole function is to obtain answers.

This analysis, together with the process of developing a benchmark and the previously discussed goals of a benchmark as well as the specifics about the MMLU, suggest that this text does not correspond to the text types described in literature. These

characteristics may have implications for the translation strategy adopted when approaching the MMLU and, consequently, on the translation techniques employed.

2.4. Translation Strategy and Techniques

Molina and Hurtado Albir (2002, p. 12) describe strategy as "the procedures (conscious or unconscious, verbal or nonverbal) used by the translator to solve problems that emerge when carrying out the translation process with a particular objective in mind". They also state that strategies are activated not only because of the difficulty of a certain unit, but also because of the translator's lack of knowledge, making the application of strategies not

The concept of strategy is relevant in the context of the MMLU in that, given that the text is not "coherent" in the sense established by De Beaugrande and Dressler's, is highly thematically heterogeneous, and it's segments are decontextualised, the translator must apply a strategy that adapts to the specific conditions of each segment.

According to Molina and Hurtado Albir (Ibid., p. 13):

> A technique is the result of a choice made by a translator, its validity will depend on various questions related to the context, the purpose of the translation, audience expectations, etc. […] we define translation techniques as procedures to analyse and classify how translation equivalence works. They have five basic characteristics:
>
> 1) They affect the result of the translation
> 2) They are classified by comparison with the original
> 3) They affect micro-units of text
> 4) They are by nature discursive and contextual
> 5) They are functional

If translation techniques are understood as a reflection of strategies, their practical execution, then the strategy adopted directly determines the techniques applied in the resolution of translation problems.

Among the translation techniques proposed by Molina and Hurtado Albir (Ibid., pp. 13-15), only the following ones were used throughout the translation process:

- Adaptation. To replace a ST cultural element with one from the target culture.
- Amplification. To introduce details that are not formulated in the ST: information, explicative paraphrasing.
- Borrowing. To take a word or expression straight from another language. It can be pure (without any change), […] or it can be naturalized (to fit the spelling rules in the TL).
- Discursive creation. To establish a temporary equivalence that is totally unpredictable out of context.
- Established equivalent. To use a term or expression recognized (by dictionaries or language in use) as an equivalent in the TL.
- Literal translation. To translate a word or an expression word for word.

2.5. Localisation and Internationalisation

The translation of a benchmark dataset into another language raises the question of localisation (often abbreviated as l10n), which means the source content is adapted to the culture of the place where the translated product is to be used, also known as locale (Anastasiou & Schäler, 2009). Jiménez-Crespo (2018) defines localisation as a "complex, technological, textual, communicative, and cognitive process by which source interactive digital texts (i.e. software, websites, and video games) undergo modifications so that they may be used in linguistic and sociocultural contexts other than those of production". Apart from translation, the linguistic aspect of localisation, it also has a technical one, which includes the adaptation of data formats and backend code, among others. This is where internationalisation (often abbreviated as i18n) comes into play: "it is the process of designing and developing a product so it can be adapted for users of different cultures and languages" (Phrase, 2025). Internationalisation aims to make the product easy to localise (Corte, 2002).

The questions of the MMLU have a very strong US-point of view, as will be further explored, and because of the structure and formulating of the questions we can confidently say that it is not a product created with internationalisation in mind, for adapting it to other cultures or locales is neither technically nor conceptually easy. However, considering the text type (described in section What Text Type does the MMLU belong to?) and the translation brief (described in section Objectives and Scope below) that determines that translation goes beyond the transfer of linguistic content, localisation is not merely applicable but a fundamental aspect of the project.

Given that localisation is applied according to the needs of the client or end user or reader (Corte, 2002), that in this case the text is aimed at LLMs, and the conditions in which this project is developed (see section Description and Goal below), the MMLU will not be localized to a specific locale, but rather subjected to localisation acts adapted to particular contextual conditions. Therefore, the use of localisation throughout the development of this project will be limited to date and time formats and units of measurement, as well as the occasionally adaptations required by the target culture or language. It is relevant to say that previous to the beginning of the translation phase, every question with a strong US-specific perspective was deleted to facilitate the process.

The effect that these changes can have on the benchmark's intention, that is, if they will alter it or have no effect over it, is unknown. It will depend on the translation techniques applied to solve each of the problems that the ST presents, which will be explored at the end of this study.

2.6. Style Guides

The following sections explore what a style guide is and its relevance, before presenting the documents that the EU Guidelines contain and focusing specifically on the Interinstitutional Style Guides.

2.6.1. Purpose

In translation, a style guide is a document that outlines the rules and standards for translating according to the client's expectations (Ghosal, 2026). Once developed, it provides a framework for translators that assures consistency in tone, terminology, formatting, and grammar, as well as aiming to preserve the original meaning and communicative intentions.

While style guides are already valuable in single-language pair translation projects, their importance increases substantially in multilingual contexts. As projects expand from one language pair to multiple target languages (TLs), the risk of stylistic inconsistency, terminological divergence, and uneven tone grows significantly (Ghosal, 2026). In this setting, style guides play a crucial role in ensuring coherence across languages by standardising linguistic conventions, register, and communicative intent, thus enabling multiple translators and teams to work in parallel while adhering to shared guidelines. Moreover, in multilingual projects, stylistic decisions that might remain implicit in monolingual workflows must be explicitly documented to prevent inconsistencies form propagating across languages (Interpro, 2024).

Not only do guidelines serve in language-related tasks, for they are also valuable when describing the actual workflow and establishing what steps a certain translation should follow. As a result, style guides become not merely supportive resources but essential tools for maintaining quality, efficiency, and consistency in multilingual translation workflows (International Organization for Standardization, 2015; TÜD SÜD, n.d.).

2.6.2. How to Build a Style Guide

In any translation and/or localization project, well-defined guidelines are necessary to ensure stylistic consistency, geographic and cultural suitability, and adherence to project requirements (Jiménez-Crespo, 2010). Style guides provide a common point of reference for all translators, particularly in workflows involving multiple translators, languages, or revision cycles, and help minimize inconsistencies.

When building a style guide, many factors must be considered. Beginning with its purpose, along with who the target audience is. This allows the translator to have a much broader picture of the goals and scope of the project. A style guide should also clarify the tone, voice, and level of formality. This is usually in line with how the brand wants to portray itself or, in the case of a benchmark, what it aims to test and how it does so.

Furthermore, grammar and writing conventions should be established, as well as terminology use (Moyer, 2001, pp. 16-17).

Style guides should also provide insight on cultural considerations (nuances, taboos, symbols…), formatting and layout of the outcome, and whether AI or NMT translation use is acceptable and which regulations should be consulted and followed (Interpro, 2024).

2.6.3. About the EU Guidelines and the Interinstitutional Style Guide

Most organizations already have guidelines developed for their translation and/or localization projects. This is the case of the European Union (EU). The ones of relevance to this project are the language-specific guidelines aimed at translators to ensure clarity, consistency, and accuracy (European Union, n.d.a). Generally, the EU presents a set of resources, some are developed by EU bodies, while others are issued by national language institutions; however, not all languages are supported by a dedicated style guide from their respective institutions.

The documents included as part of the Style Guides (European Commission, n.d.) are the following:

Language-specific style guides:	guides developed for individual TLs and addressed linguistic conventions specific to each. They are produced either by EU institutions or by the corresponding national language authorities.
Joint Handbook for the Presentation and Drafting of Acts Subject to the Ordinary Legislative Procedure:	provides practical recommendations on how legislative texts should be structures and presented when they fall under the ordinary legislative procedure. It is primarily concerned with ensuring formal consistency and clarity in EU legal acts.
Joint practical guide for persons involved in the drafting of EU legislation:	designed for those involved in drafting EU legislation in a given language. It aims to support coherence and legal precision throughout the legislative process.
DG Translation in-house style guide:	internal resource used by translators working within the European Commission that establishes stylistic and terminological conventions, reflecting established practices specific to the Commission's translation services.

Clear writing:	issued by EU institutions, these documents promote the use of plain and accessible language in official texts. Their focus lies in improving readability and ensuring that content can be understood by a broad audience.
Interinstitutional Style Guide:	uniform reference for writing and formatting across EU institutions, available in all official EU languages.

Given the overall characteristics of each document in the style guides, bearing in mind that not all of them are available for each language, the one of most interest to develop the style guide of this project is the Interinstitutional Style Guide (Publications Office of the European Union, 2022a). Mainly because it is the only one available in every EU official language that also assures aligned entries among versions, assuring some minimum contents. Furthermore, it contains information specific to conventions in every language, which no other document does and, as previously mentioned, not all languages have specific language-related guidelines created by their national institution. Therefore, in order to provide with equal resources to every language available, the Interinstitutional Style Guide is the basis of this project's style guides.

As defined by the European Union (n.d.b):

> [the Interinstitutional Style Guide] establishes writing conventions and common practices to be followed in order to present a consistent identity of the European Union institutions for citizens. Interinstitutional style guide is published in all official languages of the EU. It takes into account that while the specific character of each language has to be kept, texts in all official languages have to be comparable.

The Interinstitutional Style Guide consists of the following parts:

1. Part One. This is the Official Journal of the European Union, "the official publication (gazette) for EU legal acts, other acts and official information from EU institutions, bodies, offices and agencies" (European Union, n.d.c).
2. Part Two. This is the General Publications part, that provides guidance on presenting and drafting non-legislative EU texts (Publications Office of the European Union, n.d.a).
3. Part Three. Here, conventions common to all languages are described. Those are "linguistic rules that the European Union institutions have agreed on to ensure harmonisation in the 24 official EU languages" (Publications Office of the European Union, n.d.b).
4. Part Four. This is where each style guide focuses on the specific rules and conventions for each language, together with a list of reference works,

which complement the common conventions (Publications Office of the European Union, n.d.c).

Having established that the guidelines selected for this project were chosen precisely because of their multilingual availability, content alignment across language versions, and shared terminological resources, the same reasoning naturally shapes the extent to which they are applied; namely, only those parts that directly serve the project's translation-focused aims. With this in mind, only those parts strictly concerned with language are of relevance here. Part One and Part Two fall outside the scope of this project, for both parts are oriented towards the production and formatting of original EU documents rather than towards the linguistic and stylistic decisions faced by translators in the scope of this project, which does not involve legal or institutional EU documentation. Including them would therefore introduce guidance that, while valuable in other contexts, falls outside the scope of a workflow where the primary concern is linguistic equivalence rather than document production.

In contrast, of special interest for this project are parts three and four. Part Three indicates the common conventions on writing country names, languages and currencies, as well as their abbreviations and listing order (Publications Office of the European Union, n.d.b). Part Four guides translators through punctuation, spelling, capitalisation (upper- and lower-case usage), numbers, dates, time… But the information contained in Part Four adapts to the needs of each language, so even if most Interinstitutional Style Guides' part four contain the same information overall, there can be considerable differences (Publications Office of the European Union, n.d.c).

3. METHODOLOGY

This section outlines the methodological approach adopted to develop the style guide for the present project. It describes the objectives and scope, as well as the steps followed to analyse the source data, find the appropriate institutional resources on which to base them, and adapt them to the specific needs of the project. Ultimately, the final style guide will be described.

3.1. Project Overview

The following sections describe the project in which the present study is framed, exploring first the previous translations that have been carried out of the MMLU, before presenting the current translation brief.

3.1.1. Description and Goal

This project was proposed by the AI Unit of the Directorate General for Translation (DGT) of the European Commission in collaboration with the European Masters in Translation (EMT) network and aims to satisfy the need for a European-based benchmark for LLM evaluation. Given its leading role, the MMLU was chosen as the object of localisation. To do so, a multilingual group of master's students from the EMT member universities has been established, for the goal is to produce a human translation.

The first phase of the project involved the localisation of the MMLU into 11 European languages, process that started in early March and is due to end by 30 June 2026. In this trial run, the total of words translated for each language was dependent on the total of students available for each language pair, as well as the total of words that each committed to. Because the translation is strictly carried out by students, they too conduct the reviewing process, and all is managed through Trados Cloud.

3.1.2. Translation Brief

3.1.2.1. Previous Translation of the MMLU

Before addressing the translation brief of the project, it is fundamental to know that this dataset has already been translated in two occasions. Both translations were carried out by OpenAI. The first time, as the *GPT-4 Technical Report* created by OpenAI (2023, p. 7) states, Azure Translate (a machine translation service) was used to provide said translations, which were used to test GPT-4 in other languages aside from English and better comprehend the model's capabilities. Those languages were, in order of better to worse performance: Italian, Afrikaans, Spanish, German, French, Indonesian, Russian, Polish, Ukrainian, Greek, Latvian, Mandarin, Arabic, Turkish, Japanese, Swahili, Welsh, Korean, Icelandic, Bengali, Urdu, Nepali, Thai, Punjabi, Marathi, and Telugu. The report also covers the state of the translation, saying that "The translations are not perfect, in some cases losing subtle information which may hurt performance."

The second time, however, human translators were put in charge of the translations, therefore increasing the reliability of the output, especially in terms of accuracy, as reports the *OpenAI o1 System Card* submitted by OpenAI (2024, p. 40). The TLs were Arabic, Bengali, German, Spanish, French, Hindi, Indonesian, Italian, Japanese, Korean, Brazilian Portuguese, Swahili, Yoruba, and Simplified Chinese. After this publication, the benchmark was identified as MMMLU, *Multilingual* Massive Multitask Language Understanding. As can be observed, the MMLU was only translated into three EU official languages: German, French, and Italian. Brazilian Portuguese does not count in this context, for the variety does matter in terms of localisation. For that same reason, Spanish is also not included in the list, since the language variety used was the Hispanic American Spanish. It is worth noting that no localisation was applied, only a direct translation of the questions and answers.

A third attempt at translating the MMLU into 42 languages (Amharic, Arabic, Bengali, Chinese, Czech, Dutch, English, Filipino, French, German, Greek, Hausa, Hebrew, Hindi, Igbo, Indonesian, Italian, Japanese, Korean, Kyrgyz, Lithuanian, Malagasy, Malay, Nepali, Nyanja, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Sinhala, Somali, Shona, Spanish, Swahili, Swedish, Telugu, Turkish, Ukrainian, Vietnamese, and Yoruba) was carried out by Singh et al. (2025), which acknowledged the biases in the MMLU and, therefore, in prior attempts at translating it, and also aimed to flag content depending on whether it was “culturally-sensitive” or “culturally-agnostic”.

3.1.2.2. Current Translation Brief

Every translation responds to a specific need and set of circumstances. In other words: every translation has a purpose or *skopos*, the technical term coined by Reiss and Vermeer. The *skopos* should be established before starting the translation, for it defines how it is carried out (Schäffner, 1998), and a translation brief is a way to do so. A translation brief is a document that outlines the guidance and instructions that equip the translator with all the information needed to satisfy the client’s requirements. It includes the client information, the name and the type of project. It should also describe the purpose of the translation and the target audience (age, sex, education level, social status, etc.), as well as the specific country and region within it to which the translation is aimed. Many times, providing the translator with reference materials and glossaries will also guide them through the translation process (Vingert, 2021).

As for the translation brief of this project, it covers the following aspects:

- The text to be translated is the MMLU, a multiple-choice benchmark dataset used to evaluate and compare LLMs based on their general knowledge. The data covers 57 different topics and the level of difficulty ranges from elementary to expert.

- The ST is in English and shall be translated into your mother tongue, [specify the language depending on the translator, choose among the following: CZ, DE, EL, ES, FR, HU, IR, IT, NL, PL, PT].
- The goal is to create a target text (TT) that maintains the ST's function, that is, to keep the same level of difficulty of the questions.
- The translation is aimed at LLMs.
- It should also cover deadlines.

Previous translations of the MMLU explored in the Previous Translation of the MMLU section do not fit the requirements for this project outlined above, making them non-suitable to act as a European-based benchmark.

3.2. Objectives and Scope

The style guide developed for this project is designed to support translator through the translation process of the MMLU, while establishing a shared set of rules to ensure consistency across all languages. This is particularly important in large multilingual projects, where maintaining accuracy is essential and individual translation choices can easily diverge. By providing common guidance, the style guide helps translators approach the text consistently and apply the same criteria when making choices.

The objective, therefore, is to develop a style guide that fulfils these requisites. Achieving this first requires an understanding of the characteristics of the ST, together with the aims pursued through its translation. In this project, the defining factor shaping the style guide is the extent of localization involved, understood as the level of adaptation expected from translators. As previously established in the Localisation and Internationalisation section, localisation has been intentionally kept to minimum, since this is the first edition of the Europeanised benchmark and more extensive localization would entail significantly greater time and resource investment.

Consequently, only those questions that do not present a clearly US-specific perspective have been retained, and adaptations are limited to elements (mainly units of measurement) whose modification does not alter either the meaning of the question or the correctness of the answer. An illustrative example of this approach is provided below:

> *Mr. Fuller wants to put fencing around his rectangular-shaped yard. The width of the yard is 55 feet and the length is 75 feet. How many feet of fencing does Mr. Fuller need?*
>
> *a) 130* *b) 260* *c) 3,905* *d) 4,125*
>
> *STEM, dgt_eu_mmlu_55_en.xlsx, a77117ed-4a43-4ebf-b0d4-ad8b06c424bf__2335__test__en*

As seen in this example, by changing feet for meters and keeping the values untouched the result remains the same while also localizing it to the EU.

An illustrative example of a question where units of measurement (in this case, currencies) cannot be changed because it could alter the proposed answers, as well as the right one, is provided below:

> *As of 2020, about how much do you need to make annually to be among the richest one percent in the world?*
>
> *a) $150,000 b) $75,000 c) $35,000 d) $15,000*

Global_facts, List_of_questions_and_choices_99, question 88

This decision directly affects the development of the style guide, as it considerably reduces the level of detail required. Because the translation task does not involve the localisation of content, the guide focuses mainly on grammatical and syntactic conventions specific to each TLs. For this reason, the EU Interinstitutional Style Guide is deemed sufficient as the basis for the style guide used in this project.

In addition to addressing linguistic aspects, the style guide must also establish the methodological principles to be followed during the translation process. In this case, due to the simplicity of the translations, the process of translation will not be outlined in the style guide. Regarding the methodology of translation, it will only be made explicit the prohibition of the use of AI. This decision, made by the translation team, is based on ethical considerations and existing regulations, as previously explored.

3.3. Analysis of the Source Material

As previously discussed throughout this study, the files for translation constitute a benchmark dataset, therefore presenting a fixed structure carefully designed to test the performance of LLMs. Any alteration introduced during the translation process (such as inconsistencies or a lack of precision) may directly affect the validity of the evaluation. The dataset consists of multiple-choice questions with four possible answers, which require a high degree of accuracy when rendering the content and significantly limit the translator's freedom to reformulate, as the relationship between each question and its correct answers must be preserved at all times. The thematic diversity of the questions, together with the variation in their level of difficulty, make it necessary to maintain a neutral register, in line with the ST, and to avoid ambiguities or subjective interpretations that are not already present in the original.

Regarding localisation, as outlined above, this project requires a minimal degree of adaptation. In addition, an initial screening of the questions was carried out by the DGT in order to remove items that clearly reflected a US-centred perspective. This process reduced the volume of material to be translated and helped simplify the translation task. For this reason, the present style guide does not include guidelines related to localisation.

3.4. Identification of Relevant DGT Guidelines and Analysis

As outlined in the previous section, the Interinstitutional Style Guide provides a comprehensive set of conventions for EU publications across the 24 official EU languages. However, as established in the About the EU Guidelines and the Interinstitutional Style Guide section, only Parts Three and Four are used in this project, since Parts One and Two address aspects that do not speak to the needs of this project. The following sections outline which entries of each were selected and the reasoning behind each choice. Given the translation-focused nature of the project, the selection was guided in both cases by linguistic relevance, that is, entries directly related to grammatical and syntactic conventions, either shared between languages or particular to a few.

Part Three comprises the following chapters and entries (Publications Office of the European Union, 2022a, pp. 106-146), as described in the table below (*Table 1*):

6. Typographic instructions and revision of the text

This chapter addresses typographic instructions and presentation standards for multilingual publications.

6.1. Reading the manuscript

It addresses manuscript preparation and proofreading by the Publications Office.

6.2. Reading electronic manuscripts

It addresses electronic manuscript preparation and automates style checking.

6.3. Proof-correction marks

It addresses the review and correction of first proofs by Publications Office proofreaders.

6.4. Word-processor punctuation marks and spacing

It establishes a common standard on how to apply typographic codes in the EU's multilingual environment.

6.5. Punctuation in figures

It establishes the use of decimal and thousand separators, with specific guidance on budgetary figures.

7. Countries, languages, currencies and their abbreviations

It covers harmonised conventions for referring to country, language and currencies.

7.1. Countries

It addresses the standardised designation and abbreviation of EU member, candidate, and third countries.

7.1.1. Designations and abbreviations to use

It establishes the rules that must be followed when writing and abbreviating the names of the Member States of the EU.

7.1.2. Country listing order

It addresses the listing order of countries in text and tables.

7.2. Languages

It covers listing conventions for language versions as opposed to language names in text.

7.2.1. Order of language versions and ISO codes (multilingual texts)

It established the order in which language versions should appear in multilingual publications.

7.2.2. Listing language in text (unilingual texts)

It addresses the ordering of language names in unilingual publications.

7.2.3. Indication of languages published

It addresses how language versions should be presented when listed in catalogue or similar contexts.

7.2.4. Rules governing the languages of the institutions

It addresses the regulatory framework governing the official languages of EU institutions.

7.3. Currencies

It covers the standardised use and notion of currencies across EU publications.

7.3.1. The euro and the cent

It addresses the official name and notation of the euro and the correct use of the term “euro area”.

7.3.2. Order of the currencies and ISO codes

It addresses ISO currency codes and the listing order of Member State currencies.

7.3.3. Rules for expressing monetary units

It establishes how to properly indicate monetary units depending on the case.

8. Footnotes and their references

It covers footnote conventions, including their placement, numbering, and treatment of repeated notes.

8.1. Footnote references

It covers the symbols and figures used as footnote references, which are consistent across all language versions.

8.2. Order of footnotes

It covers the consistent ordering of footnotes across language versions.

9. Other conventions

It covers additional conventions such as addresses, electronic addresses, telephone numbers, citations, and administrative structure of the European Union.

9.1. Addresses

It covers the conventions for writing addresses in EU publications, including rules for unilingual, multilingual, and Member State addresses.

9.2. Electronic addresses

It covers conventions for writing electronic addresses.

9.3. Telephone numbers

It covers conventions for formatting of telephone numbers in accordance with ITU recommendation.

9.4. Citation of works and references to the Official Journal

It covers citation conventions for bibliographies.

9.5. Administrative structure of the European Union: official titles and listing order

It covers the standardised designation of the EU's institutional and administrative bodies, including agencies and interinstitutional services.

9.6. Commission directorates-general and services: official titles

It covers the internal coding system used to identify EU institutions and bodies.

Table 1. Annotated index of Parth Three of the Interinstitutional Style Guide.

Within Part Three, the selection below alone was deemed relevant due to its linguistic nature, as the table shows (*Table 2*):

6.4. Word-processor punctuation marks and spacing
7.1.1. Designations and abbreviations to use
7.2.1. Order of language versions and ISO codes (multilingual texts)
7.3.3. Rules for expressing monetary units

Table 2. Selection of entries from Part Three of the Interinstitutional Style Guide.

The remaining entries deal with document production procedures and administrative conventions that, while relevant to EU publishing practices, are not concerned with the linguistic aspects central to this project.

As for Part Four, almost every aspect covered by the Interinstitutional Style Guide is relevant to the current project, therefore they were almost all included in the final style guide. The sole criterion for exclusion was whether a given entry addressed conventions

specific to a single language rather than shared across several. Given the number or languages involved in the project, incorporating language-specific entries for each would have made the style guide overly complex and difficult to use. Translators were nonetheless encouraged to refer to the complete language-specific guidelines at their own discretion, as they may contain additional relevant information.

In the case of the Spanish guidelines (Publications Office of the European Union, 2022b, pp. 143-185), all chapters and entries of the Interinstitutional Style Guide's Part Four were included, as every aspect covered proved relevant and was shared across the other languages. These are explored in Spanish in the table below (*Table 3*):

10. Presentación formal del texto
10.1. Puntuación
10.1.1. Coma
10.1.2. Punto
10.1.3. Punto y coma
10.1.4. Dos puntos
10.1.5. Puntos suspensivos
10.1.6. Paréntesis
10.1.7. Corchetes
10.1.8. Raya y semimenos
10.1.9. Guion
10.1.10. Barra
10.1.11. Interrogación y exclamación
10.1.12. Comillas
10.1.13. Acentuación
10.1.14. Particularidades tipográficas
10.2. Mayúsculas y minúsculas
10.2.1. Entidades
10.2.2. Denominaciones relacionadas con la actividad jurídica o intelectual
10.2.3. Publicaciones y obras de creación
10.2.4. Personas
10.2.5. Lugares
10.2.6. Otros
10.3 Números
10.3.1. Números cardinales
10.3.2. Números ordinales
10.3.3. Números romanos
10.4. Abreviaturas

10.5. Siglas y acrónimos
11. Obras de consulta

Table 3. Selected chapters and entries from Part Four of the Spanish version of the Interinstitutional Style Guide.

It is important to highlight that these style guides do not contain localisation-specific information. While this poses no problem to this version of the style guide because the project entails almost no localisation, it could be potentially problematic in future versions of the present style guide, if the project were to be taken further.

3.5. Design and Justification of the Project Style Guide

In the first page of the style guide, a small introduction is presented. The source dataset is not described (as well as the target audience) because the translators are already familiar with it, but it is explicitly mentioned. Then, a small description of the style guide is presented, so translators will know what it contains, as well as an overview of the entries included with direct links to them.

Given that the project style guide is solely based on the Interinstitutional Style Guide of the EU, and knowing that it entails almost no localization, the optimal design is to create a table, where each column represents a language and each row represents an aspect of the style guide. This makes it easier for translators to quickly search through the style guide and easily find each entry.

As for how information is shown, the Interinstitutional Style Guide has an online page where individuals can browse through it. For that reason, in the project style guide a link to the respective entry in the online version is provided, as shows the example below (*Table 4*).

	BG	ES	CS	DE	ET	EL	EN	FR	GA	IT	LV	HU	NL	PL	PT	SL
1. Country designations and abbreviations	7.1.1. Наименования и съкращения, които сле	7.1.1. Denominaciones y siglas preceptivas	7.1.1 Názvy a zkratky	7.1.1. Zuverwendende Länderbezeichnungen und	7.1.1. Kasutusel olevad nimetused ja lühendid	7.1.1. Ονομασίες και συντομογραφίες	7.1.1. Designations and abbreviations to use	7.1.1. Dénominations et sigles à utiliser	7.1.1. Ainmneacha agus giorrúcháin a úsáidtear	7.1.1. Denominazione sigle da utilizzare	7.1.1. Izmantojamie apzīmējumi un abreviatūras	7.1.1. Megnevezések és rövidítések	7.1.1. Benaming en te gebruiken afkortingen	7.1.1. Stosowane nazwy i skróty	7.1.1. Denominações e siglas a utilizar	7.1.1 Imenain kratice

	два да бъдат използвани			Kürzel												

Table 4. First row of project's style guide.

The sections included are:

1. Country designations and abbreviations
2. Languages
3. Rules for expressing monetary units
4. Punctuation
5. Word-processor punctuation marks and spacing
6. Capitalisation (upper and lower case)
7. Numbers, dates and time
8. Abbreviations
9. Acronyms and/or initialisms
10. Italics
11. Verbs
12. Spelling
13. Prepositions
14. Use of articles
15. Miscellaneous

Aside from 1., 2., 3., and 5., which belong to Part Three and are therefore common to all languages, in order to create a complete style guide, every Part Four of each language was examined and its content evaluated and then standardised so as to fit as many entries of each language into one of the categories. Entries that were not related to any of them were put in the miscellaneous section. Furthermore, a link to every Interinstitutional Style Guide's home page for every language was added at the end of the document, so that translators could access it freely and consult it further if needed.

In this first edition, the style guide has deliberately been kept concise, resulting in a functional scope that is limited to strictly linguistic aspects. The level of detail provided is sufficient to ensure consistency among translators working into the same language, while a higher degree of specificity was deemed unnecessary, as the ST does not present major translation challenges. For the same reason, the EU Interinstitutional Style Guide was considered an appropriate reference, as it offers the relevant guidance for this project and situates it withing established EU conventions, which is particularly suitable given that the project itself is promoted by the European Union.

The definitive style guide is attached as the APPENDIX.

3.6. Translation Problems Beyond the Current Style Guide

Throughout the translation process, a number of translation problems inherent to the ST were identified. Since they had not been anticipated in the original style guide, they point to the need to address such issues when developing guidelines for similar projects in the future. These problems are examined in relation to the specific purpose of the translation task: producing a TT intended for interaction with an LLM. From a functional perspective, the ST is designed to assess the ability of an LLM to solve different types of problems while maintaining a certain level of difficulty, a characteristic that the TT must preserve. This requirement directly shapes the choice of translation techniques throughout.

The translation problems identified are analysed systematically. The following table (*Table 4*) presents the template used to describe them, detailing the possible courses of action and their respective implications.

1. Problem Identification
    1.1. Problem Source. Here the segment in question is presented.
    1.2. Problem Type. Classification of the problem presented by the segment in relation to the function of the text, according to the following categories:
        1.2.1. If it affects ST integrity
            1.2.1.1. Formal level
                - **Typo**: the segment contains a typographical error.
                - **Misspelling**: the segment contains misspellings affecting one or more words.
            1.2.1.2. Content level
                - **Missing segment**: the segment is missing a portion of text, rendering it incomprehensible.
                - **Lack of reference with real world reality**: the segment contains factually inaccurate information.
        1.2.2. If it affects content reliability
            - **Factually outdated information**: the segment contains information that is outdated or no longer in common use.
            - **Culturally insensitive content**: the segment contains content considered culturally insensitive.
            - **Cultural bias**: the segment reflects a US-centric perspective.
        1.2.3. If it affects TT comprehension
            - **Concept requiring expansion**: the segment contains a concept that requires further explanation to be fully understood in the TL.

- **Wording and disambiguation**: the segment contains wording that requires disambiguation in order to be accurately transferred to the TL.

1.3. Action with Regard to Textual Function. Based on the function of the text, aiming at not altering the level of difficulty:

- **The problem must be preserved**. These are cases where the problem does affect how the segment is evaluated, making it essential to retain the error in the TL so as to replicate the conditions of the ST.
- **The problem does not need to be preserved**. The problem does not affect segment evaluation, as it either relates to a peripheral aspect of the segment or affects an element that is not essential to the comprehension of the question, therefore fixing it will not have a negative impact on the function of the text, and the original conditions of the segment do not need to be replicated.

2. Solution *n*

2.1. Description of the Solution and Explanation. Three possible approaches are proposed to address the problem, depending on the nature of the error and its impact on the translation. When making decisions, the guiding question is how best to honour the intention of the text.

- **Preserving the original conditions without localisation or cultural adaptation** applies to cases where the error affects the difficulty level of the segment, but the problematic element does not require localisation, for instance, a proper noun or the title of a work.
- **Replicating the original conditions through localisation** applies to cases where translating without localisation increases the level of difficulty in the TL. This occurs when the segment draws heavily on cultural elements specific to a country other than that of the target culture, or when the aim is to propose a problem equivalent to the original but grounded in the culture of the TL.
- **Eliminating the problem** is applicable in cases where the error does not affect the difficulty level of the segment and therefore does not need to be retained. This applies primarily to problems related to ST integrity, in particular typographical and spelling errors.

2.2. Translation Technique Applied. According to the translation techniques proposed by Molina and Hurtado Albir and (2002), the following may be applied:

2.2.1. Expansion: the TT contains more information than the ST.

- **Adaptation**
- **Amplification**

- **Discursive Creation**

2.2.2. Preservation: the TT conveys the same information as the ST, though it may be presented differently.

- **Borrowing**
- **Established Equivalent**
- **Literal translation**

2.3. Output regarding Level of Difficulty.

- **Preserved**. The solution preserves the intended level of difficulty of the segment, in accordance with the functional requirements of the task.
- **Reduced**. The solution results in a target segment that is easier to process than the original, for instance by making explicit information or by resolving an ambiguity present in the ST.
- **Increased**. The solution results in a target segment that is harder to process than the original, for instance by retaining a cultural reference without further explanation.

2.4. Translation output: resulting translation of the segment in the TL (ES), reflecting the solution applied.

Table 5. Problem analysis template.

This template is applied to the each of the recurring translation problem types identified, as detailed below.

PROBLEM No. 1

1. Problem Identification
   1.1. Problem Source

> Q: […] All **lettres de cachet** shall be abolished forever, subject to certain modifications which the State General may see fit to impose.
>
> *List of questions and choices 50, question 49*

   1.2. Problem Type: **Concept requiring expansion**.
   1.3. Action with Regard to Textual Function: **The problem does not have to be preserved**.
2. Solution 1
   2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**
   2.2. Translation Technique Applied: **Borrowing.**
   2.3. Output regarding Level of Difficulty: **Increased.**
   2.4. Solution: ***lettres de cachet.***
3. Solution 2

3.1. Description of the Solution and Explanation: **Replicate the original conditions through localisation.**

3.2. Translation Technique Applied: **Borrowing + Amplification.**

3.3. Output regarding Level of Difficulty: **Preserved.**

3.4. Solution: ***lettres de cachet*** **(Real Orden Reservada).**

PROBLEM No. 2

1. Problem Identification

1.1. Problem Source

> Q: Which of the following best describes the practice of placing the blame for the problems of a society on one particular group?
>
> Option 1: Ethnocentrism
> Option 2: Stereotyping
> Option 3: **Scapegoating**
> Option 4: Institutionalization
>
> *List of questions and choices 99-sociology, question 78*

1.2. Problem Type: **Concept requiring expansion.**

1.3. Action with Regard to Textual Function: **The problem does not have to be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**

2.2. Translation Technique Applied: **Borrowing.**

2.3. Output regarding Level of Difficulty: **Increased.**

2.4. Solution: ***Scapegoating.***

3. Solution 2

3.1. Description of the Solution and Explanation: **Replicate the original conditions through localisation.**

3.2. Translation Technique Applied: **Amplification.**

3.3. Output regarding Level of Difficulty: **Preserved.**

3.4. Solution: ***Scapegoating*** **o la utilización de alguien como chivo expiatorio.**

PROBLEM No. 3

1. Problem Identification

1.1. Problem Source

> Q: This question refers to the following information.
> The following excerpt is from a speech by **Vladimir I. Lenin**. [sic]*
> […]

List of questions and choices 126, question 121

**TN: This speech was not done by Lenin, but by Stalin.*

1.2. Problem Type: **Lack of reference with real world reality.**

1.3. Action with Regard to Textual Function: **The problem must be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**

2.2. Translation Technique Applied: **Literal translation.**

2.3. Output regarding Level of Difficulty: **Preserved.**

2.4. Solution: **Vladimir I. Lenin.**

3. Solution 2

3.1. Description of the Solution and Explanation: **Eliminate the problem.**

3.2. Translation Technique Applied: **Discursive creation.**

3.3. Output regarding Level of Difficulty: **Reduced[1].**

3.4. Solution: **Iósif Stalin.**

PROBLEM No. 4

1. Problem Identification

1.1. Problem Source

Q: The global HIV/AIDS death rate I highest for which of these age groups?

Option 1: Under 5
Option 2: **14-May**
Option 3: 15-49
Option 4: over 50

List of questions and choices 99-global facts, question 81

1.2. Problem Type: **Typo.**

1.3. Action with Regard to Textual Function: **The problem does not have to be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**

2.2. Translation Technique Applied: **Literal translation.**

2.3. Output regarding Level of Difficulty: **Increased.**

[1] In this case, it is not possible to determine with certainty whether the level of difficulty increases, decreases, or remains unaltered, as it is unknown whether the error was intentional and, consequently, whether it was intended to affect the level of difficulty.

2.4. Solution: **14-Mayo.**

3. Solution 2
   3.1. Description of the Solution and Explanation: **Eliminate the problem.**
   3.2. Translation Technique Applied: **Discursive creation.**
   3.3. Output regarding Level of Difficulty: **Preserved.**
   3.4. Solution: **5-14.**

PROBLEM No. 5

1. Problem Identification
   1.1. Problem Source

> Q: Which of the following mental disorders **is most prevalent in the world**?
>
> Option 1: Bipolar disorder
> Option 2: Schizophrenia
> Option 3: Alcohol use disorders
> Option 4: Anxiety disorders
>
> *List of questions and choices 99-global facts, question 97*

   1.2. Problem Type: **Factually outdated information.**
   1.3. Action with Regard to Textual Function: **The problem does not have to be preserved.**

2. Solution 1
   2.1. Description of the Solution and Explanation: **Preserving the original conditions without localisation or cultural adaptation.**
   2.2. Translation Technique Applied: **Literal translation.**
   2.3. Output regarding Level of Difficulty: **Increased.**
   2.4. Solution: **¿Cuál de los siguientes trastornos mentales es el más frecuente en el mundo?**

3. Solution 2
   3.1. Description of the Solution and Explanation: **Replicating the original conditions through localisation.**
   3.2. Translation Technique Applied: **Amplification.**
   3.3. Output regarding Level of Difficulty: **Preserved.**
   3.4. Solution: **¿Cuál de los siguientes trastornos mentales era el más frecuente en el mundo en el año 2021?**

PROBLEM No. 6

1. Problem Identification
   1.1. Problem Source

Q: As of 2019, about what percentage of Italians say it is very important to have free media in **our country** without government/state censorship?

Option 1: 41%
Option 2: 56%
Option 3: 71%
Option 4: 86%

List of questions and choices 99-global facts, question 75

1.2. Problem Type: **Cultural bias.**

1.3. Action with Regard to Textual Function: **The problem does not need to be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**

2.2. Translation Technique Applied: **Literal translation.**

2.3. Output regarding Level of Difficulty: **Increased.**

2.4. Solution: **En 2019, ¿qué porcentaje aproximado de italianos afirman que es muy importante contar con medios de comunicación libres en nuestro país, sin censura gubernamental o estatal?**

3. Solution 2

3.1. Description of the Solution and Explanation: **Replicate the original conditions through localisation.**

3.2. Translation Technique Applied: **Amplification.**

3.3. Output regarding Level of Difficulty: **Preserved.**

3.4. Solution: **En 2019, ¿qué porcentaje aproximado de italianos afirman que es muy importante contar con medios de comunicación libres en Estados Unidos, sin censura gubernamental o estatal?**

PROBLEM No. 7

1. Problem Identification

1.1. Problem Source

Q: Predict the hyperfine value for the EPR spectrum of fully **deuteriated** benzene radical anion C6D6•-.

Option 1: 0.375 mT
Option 2: 3.75 G
Option 3:2.35 mT
Option 4:0.58 G

List of questions and choices 83, question 41

1.2. Problem Type: **Misspelling.**

1.3. Action with Regard to Textual Function: **The problem does not have to be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Replicate the original conditions through localisation.**

2.2. Translation Technique Applied: **Discursive creation.**

2.3. Output regarding Level of Difficulty: **Preserved.**

2.4. Solution: **Prediga el valor hiperfino para el espectro de EPR del anión radical de benceno completamente deuteriado C6D6•-.**

3. Solution 2

3.1. Description of the Solution and Explanation: **Eliminate the problem.**

3.2. Translation Technique Applied: **No specific technique is required; the technique is determined by the translation strategy applied.**

3.3. Output regarding Level of Difficulty: **Preserved.**

3.4. Solution: **Prediga el valor hiperfino para el espectro de EPR del anión radical de benceno completamente deuterado C6D6•-**.

PROBLEM No. 8

1. Problem Identification

1.1. Problem Source

Q: In 1943 how many Americans **believed in the Holocaust**?

Option 1: 30%
Option 2: 50%
Option 3: 70%
Option 4: 90%

List of questions and choices 99-global facts, question 89

1.2. Problem Type: **Culturally insensitive content.**

1.3. Action with Regard to Textual Function: **The problem does not have to be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**

2.2. Translation Technique Applied: **Literal translation.**

2.3. Output regarding Level of Difficulty: **Preserved.**

2.4. Solution: **En 1943, ¿cuántos estadounidenses creían en el Holocausto?**

3. Solution 2

3.1. Description of the Solution and Explanation: **Eliminate the problem.**

3.2. Translation Technique Applied: **Adaptation.**

3.3. Output regarding Level of Difficulty: **Preserved.**

3.4. Solution: **En 1943, ¿cuántos estadounidenses pensaban que el Holocausto realmente estaba ocurriendo?**

PROBLEM No. 9

1. Problem Identification

1.1. Problem Source

> Q: […] Should disputes or conflicts arise between the High Contracting Parties* shall participate in any grouping of Powers whatsoever that is directly or indirectly aimed at the other party.
>
> Option 1: Article I
> Option 2: Article II
> Option 3: Article III
> Option 4: Article IV
>
> *List of questions and choices 141, question 140*
>
> *NT: There is an elided fragment.

1.2. Problem Type: **Missing segment.**

1.3. Action with Regard to Textual Function: **The problem does not need to be preserved.**

2. Solution 1

2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**

2.2. Translation Technique Applied: **Literal translation.**

2.3. Output regarding Level of Difficulty: **Preserved.**

2.4. Solution: **[…] Si surgieran disputas o conflictos entre las Altas Partes Contratantes participará en cualquier agrupación de potencias de cualquier tipo que esté dirigida directa o indirectamente contra la otra parte.**

3. Solution 2

3.1. Description of the Solution and Explanation: **Eliminate the problem.**

3.2. Translation Technique Applied: **Amplification.**

3.3. Output regarding Level of Difficulty: **Preserved.**

3.4. Solution: **[…] Si surgieran disputas o conflictos entre las Altas Partes Contratantes sobre problemas de una u otra índole, ambas partes**

participarán en cualquier agrupación de potencias de cualquier tipo que esté dirigida directa o indirectamente contra la otra parte.

PROBLEM No. 10

1. Problem Identification
   1.1. Problem Source

> Q: Allen's (2001) study of **nurses** showed that:
>
> Option 1: **nurses** perform a great deal of emotional labour for which they are not paid
> Option 2: the policies of Project 2000 had the effect of de-professionalising the career of nursing
> Option 3: more men were moving into nursing, which challenges some of the prevailing stereotypes about the occupation
> Option 4: the role of the **nurse** was ambiguous and had to be constantly redefined in the context of everyday interactions with other staff and patients
>
> *List of questions and choices 99-sociology, question 77*

   1.2. Problem Type: **Wording and disambiguation.**
   1.3. Action with Regard to Textual Function: **The problem does not need to be preserved.**
2. Solution 1
   2.1. Description of the Solution and Explanation: **Preserve the original conditions without localisation or cultural adaptation.**
   2.2. Translation Technique Applied: **Literal translation.**
   2.3. Output regarding Level of Difficulty: **Preserved.**
   2.4. Solution: **El estudio de Allen (2001) sobre las enfermeras mostró que:**
   Opción 1: **las enfermeras ejercen una labor emocional por la que no se les retribuye**
   Opción 4: **el papel de las enfermeras era ambiguo, por lo que tenía que ser constantemente redefinido en el contexto de las interacciones diarias tanto con otros miembros del equipo médico como con los pacientes**
3. Solution 2
   3.1. Description of the Solution and Explanation: **Eliminate the problem.**
   3.2. Translation Technique Applied: **Amplification.**
   3.3. Output regarding Level of Difficulty: **Preserved.**
   3.4. Solution: **El estudio de Allen (2001) sobre las personas que se dedican a la enfermería mostró que:**
   Opción 1: **los y las enfermeras ejercen una labor emocional por la que no se les retribuye**

Opción 4: **el papel de los y las enfermeras era ambiguo, por lo que tenía que ser constantemente redefinido en el contexto de las interacciones diarias tanto con otros miembros del equipo médico como con los pacientes**

4. DISCUSSION

The following section discusses the findings of this project. Starting with the style guide developed, it can confidently be said that it was comprehensive within the parameters for which it was designed; however, those parameters did not fully account for the specific nature of the MMLU as a text type or for the particular function of this translation assignment, rendering the guide insufficient. As a result, it failed to provide with enough insight to follow an established strategy or apply certain translation techniques depending on the problem type. This explains the discrepancies between translation solutions and the absence of fixed criteria, as the guide was originally designed for a different type of text that did not include cultural references. Although a selection phase prior to the translation process had been carried out in order to delete culturally charged question, it was later observed that many other questions presented a more subtle but equally evident cultural bias. The guide was therefore conceived for a different set of problems.

Furthermore, the fact that the text was being translated for a machine rather than a human was not clearly identified form the outset, which meant that the function of the translations was not adequately defined. The guide tackled aspects like measurement units or upper- and lower-case usage, but it failed to cover what to do when cultural problems appear. These gaps in the style guide became apparent one the translation process began and the functional shift implied by translating for a machine (rather than a human reader) became evident.

4.1. Assessment of Translation Solutions

As for the problems documented in the previous section, they collectively confirm that the limitations of the style guide, which was not adapted to the characteristics of the ST, is responsible for the issues encountered. The fragmented, decontextualised, and thematically heterogeneous nature of the text constitutes the primary source of translation problems (with the exception of typographical and spelling issues) since, has these parameters been uniform throughout, the translator would have been able to make informed decisions based on context and a more clearly defined translation brief. However, it is the style guide that has to explored said characteristics in order to provide with a translation strategy and techniques that go accordingly, and it failed to address them adequately, which in turn allowed inherent textual features to become actual translation problems.

The problems encountered can be grouped into broader categories, each of which calls for a general course of action, as outlined below. These approaches can be extrapolated to equivalent translation problems.

The problems that, corrected/localised or not, do not pose a variation of the level of difficulty of the segment (Problems 7, 8, 9, and 10) are based on grammar (misspellings),

wording and missing information. In those cases, keeping the error is aimless, for it is assumed that the model will infer information and correct mistakes. Therefore, in order not to build a benchmark with futile errors, the problematic segments should be corrected through the application of translation techniques. It is worth mentioning that in Problems 8 and 10 localisation means alienating them with European values, which is in line with the general aim of the project.

As for the problems that do suppose a change in the level of difficulty (Problems 1, 2, 3, 4, 5, and 6), the general idea when making a choice on what to do is to maintain the original level of difficulty. Therefore, most of them could be solved through the localisation of the problematic segment. The degree to which localisation is applied depends on the aim of the project, that is, whether it is being able to use the MMLU in other countries aside from the US or domesticating it so as to create an EU-centred benchmark.

The following table (*Table 5*) provides an overview of the problems identified and the solutions proposed for each category in the project, as well as the author's assessment of each solution in relation to the overarching objectives for the project.

Problem	**Problem Category**	**Solution Adopted**	**Critical Assessment**	**Justification**
Problem 1	Concept requiring expansion	Replicate through localisation	✓	Given that the term *lettre de cachet* is the one used in English to refer to the concept, the term itself poses no additional problem. Therefore, in the TL, the term with which the language refers to said concept (the established equivalent) should be used in order to maintain the level of difficulty of the segment. Another possible solution would be the one adopted in the project, that is to keep the borrowing and add a brief explanation or the established equivalent in the TL in brackets.

Problem 2	Concept requiring expansion	Replicate through localisation	✓	Given the ability of the English language to generate words that summarise a whole idea, and given that other languages do not have that ability, in order to maintain the same answer structure in all four options, it is acceptable to keep the English word as a foreignism and add a brief explanation of what it means in the TL.
Problem 3	Lack of reference with real world reality	Preserve without cultural adaptation	✓ / ✗	The acceptability of this decision varies depending on whether the mistake was made on purpose or not. If it was an intended added difficulty, then the correct decision would be to preserve the error. If the segment was not intended to have this added level of difficulty, then the segment could be corrected. However, it can also be argued that even if the problem was unintentional, it is there and affects the output of the model, therefore by eliminating it the benchmark developed would not be totally equivalent to the MMLU.

				As a counterpart, if it is unintentional, then the level of difficulty of this question is not completely taken into account, as the additional difficulty that the error introduces contributes to it. Following this line of reasoning, it is pointless from the perspective of difficulty to preserve the error, given that it adds a challenge that was never intended to be part of what the question tests.
Problem 4	Typo	Replicate through localisation	✓	The error irremediably affects the level of difficulty, though in a different way in English than in other languages. In the ST, the error most likely originates when Excel interpreted the age range "5-14" as a date, automatically replacing it with "14-May". However, this substitution would not have occurred in other languages, as Excel's date recognition is locale-dependent. Furthermore, the error does not affect what the item is designed to evaluate (namely, the model's knowledge with respect to the question

				itself). Therefore, if it were to be preserved, it would raise the difficult level compared to the original, while if the problem is corrected, the ST and TT would have more comparable difficulty levels.
Problem 5	Factually outdated information	Preserve without localisation	✘	The question tests the model's knowledge of global facts, but without specifying the year to which it refers to (presumably the year it was created), the model may infer that it refers to the specific year the benchmark is used, and then the answers may be outdated. Therefore, by adding the specific year the question refers to, the answers will always be right.
Problem 6	Cultural bias	Preserve without localisation	✘	This question presents a strong US-centric perspective and could not be used in other countries because it would not make sense, for the answer is tailored to the US. Therefore, by substituting "our country" by "the US", the question is kept valid.
Problem 7	Misspelling	Eliminate the problem	✓	The problem does not have a significant impact

				on the difficulty level, therefore it does not require the translators to reproduce it.
Problem 8	Culturally insensitive content	Eliminate the problem	✘	Although the use of the term "believe" may seem problematic in this context (given the Holocaust was indeed taking place in 1943), the question refers to the political context of the time, in which this was not yet a matter of established fact but a political position. The term is therefore considered appropriate within its historical context.
Problem 9	Missing segment	Eliminate the problem	✓	The problem does not have a significant impact on the difficulty level, therefore it does not require the translators to reproduce it.
Problem 10	Wording and disambiguation	Eliminate the problem	✓	The problem does not have a significant impact on the difficulty level, therefore it does not require the translators to reproduce it.

Table 6. Translation problem analysis and justification of author's solutions.

Because the analysis has been qualitative rather than quantitative, it is not possible to determine the exact frequency with which translators encountered each type of problem. However, Problem 6 can be identified with confidence as the most problematic, since its strong US-perspective runs counter to the overarching objective of the project, which is to develop a benchmark from a European standpoint. In this case, and with regard

to any item reflecting a distinctive American point of view, the solution varies depending on the intended outcome: if the priority is to retain the greatest possible number of questions from the original, replacing “our country” with “the US” would be enough, thereby specifying the country so that the answer remains valid; if a greater degree of domestication is sought, a new question that fulfils the same evaluative function as the original would need to be generated, but reframed from a Europe perspective.

4.2. Recommendations for Future Editions

Each of these solutions has an impact on the workflow, as well as on the style guide. For those problems that do not affect the level of difficulty of the question, it is recommended that an explanation be added to the style guide indicating that all spelling or typographical errors should be corrected in the translation. Translators could also be informed on how to handle this type of problem both in the briefing and in the informal meetings held prior to the translation phase.

For problems that do affect the level of difficulty of the question, it is recommended that a pre-editing process be carried out on the relevant segment in order to adapt its content or apply localisation. This would allow the items to be aligned with a European perspective while facilitating the translation process and ensuring that the same translation techniques are applied consistently across all TLs. If pre-editing cannot be applied, then the style guide should explore each type of problem and provide strict guidelines on how to solve them.

The results obtained from the translation process indicate, therefore, that the briefing should have defined a clear protocol for dealing with segments with a strong US-perspective, as well as segments that lack sufficient context for decision-making, those that include information with no reference with real world reality, and those problems inherent from the ST, such as typographical and spelling issues.

As for the style guide, it should have provided the translators with guidelines for managing the ST errors that affect the level of difficulty. That includes covering explicit instructions on how to deal with decontextualised segments in terms of disambiguating or expanding concepts. It should have also dedicated a section to localisation decisions, specifying when to adapt culturally specific content and how to do so (if the project includes the localisation of segments as part of the translation process).

And lastly, they were a major indicator that pre-editing should have been applied to those culturally biased segments and, if the localisation process was not expected to be carried out by the translators, then localisation and correction of errors should have also been part of the pre-editing process.

In addition to the above, it is strongly recommended that, prior to the translation process, a preliminary study be conducted on the main problem types in order to determine whether and to what extent each one affects the LLM’s evaluation capacity.

5. CONCLUSION

All of the results presented in this study show that the translation aimed at machines that will have the goal of evaluating LLMs (but that also need to be understood by humans) suppose a new textual genre. This has implications for the briefing as well as for the function of the text.

The MMLU was designed from an American perspective that shapes not only the content of the segments but also the way in which questions are framed. This means that adapting it to reflect a European perspective goes beyond a straightforward translation process. Furthermore, not all segments are equally adaptable: while some can be localised without compromising their evaluation function, others are so deeply embedded in a specific cultural context that any adaptation would inevitably alter what they are testing. Following the development of this project, it can be concluded that while it is possible to adapt a benchmark created in a specific sociocultural context and reframe it into another, this cannot be achieved by merely translating its questions, as culture is crucial in LLM evaluation, particularly in the creation of a benchmark with a European perspective. Such a task requires an in-depth understanding of what each segment is testing, technical resources for properly localisation, and, crucially, time both for the technical work and for the translation itself.

The following sections revisit the objectives of the study, answer the research questions, and explore future lines of research.

5.1. Revisiting Research Questions and Objectives

Revisiting the research questions outlined at the beginning of this study and aiming to answer them based on the results described in 4. Discussion:

1. *Does the decontextualised nature of the MMLU generate different translation problems to those found in conventional translation projects?*

As can be observed in the Translation Problems Beyond the Current Style Guide section, as a result of the segments lacking context, the translator is left with the responsibility of making translation decisions without enough information, for each segment acts as an independent unit. This can lead to ambiguities, which in a conventional project could be solved through the context. Therefore, it has been observed that the textual typology of the MMLU does generate translation problems that differ from those found in conventional translation assignments. Moreover, this also has an impact on the translation techniques that can be applied, which differ from conventional ones in that they were not originally conceived for a scenario where the target audience is an LLM rather than a human reader.

2. *Can a style guide as conventionally defined in translation studies be used to tackle assignments such as the translation of the MMLU?*

This question is closely related to the type of text of the MMLU and to whether it is comparable to any other text type commonly described in conventional translation assignments. The style guide for this project was developed following the principles that typically govern the creation of guidelines, as explored in the How to Build a Style Guide section. However, as has been demonstrated throughout this study, the guide proved insufficient to manage the full scope of the assignment. The main reason is that this text does not fit into conventional text typologies, and therefore its function differs from that of standard translation assignments. Following the completion of this study, it has been observed that specifications for how to handle cultural references should have been included, along with provisions for issues that go beyond what is typically encountered, as the assignment extends further than the mere transfer of linguistic elements. Consequently, the MMLU cannot be adequately translated using a conventional style guide within the framework of this project.

3. *To what extent is the MMLU translatable without compromising the difficulty level and evaluative purpose for which it was originally designed?*

Applying the right translation techniques, translating the MMLU while maintain its difficulty level and evaluative purpose is achievable. In this first attempt it may not have been fully accomplished (not due to the lack of effort, but because the complexity level of the task was only fully understood once the process was already underway). In. that sense, this first edition has served as a valuable learning opportunity: the challenges encounters, the problems identified, and the solutions proposed throughout this project have generated a body of knowledge that can and should be applied in future editions, where the conditions will be better defined and the objectives more attainable.

4. *How does the nature of the MMLU redefine the role of the translator and the decision-making process?*

Throughout the project, translators have had a great deal of responsibility of not only translating but being one of the first to face the MMLU, with everything that entails. They encountered both fragments that did not seem problematic at first but ended up posing an issue to the benchmark's function, and other fragments that were clearly an issue and required an action beyond mere translation. They acted as translators, but also as pre-editors and reviewers. Therefore, in this first edition of the project, the translator role did come with some decision-making (at least, at first). Those decisions had to be discussed with the rest of the translation team as well as the project managers, and a final answer was given either by the project managers or the AI Unit of the DGT, establishing a reference point for the rest of the team. However, the decision-making "freedom" that translators had throughout this project came about only because every obstacle the

MMLU posed was encountered for the first time while translating and was not contemplated by the style guide, which did not offer strict rules on how to handle certain problems. In future editions, when the content is better explored, the localisation strategy is clear, and the text has been pre-edited, the translator will be able to face it as if it were an average project.

At the beginning of the study, four specific objectives that the project wanted to accomplish were established.

The first objective was met, which was *to develop a theoretical understanding of what benchmarks are, what they measure, and how they do so in order to understand the textual function of the text*. As well as the second, *to develop a style guide based on official EU linguistic standards*, which was done by using the Interinstitutional Style Guides of the EU.

The third objective of the project was *to identify to which extent the style guide met the needs of translators* (that is, how adjusted it was to cover the specific needs of the text type). This aspect proved more significant than initially anticipated, as the full scope of the benchmark and its implications for the translation process were not entirely comprehended. Though there was an initial understanding of how the text type and its function would affect it, it was only during the translation process itself that their whole implications became clear. Either way, the translation problem analysis served as a first exploration of the different problem types the MMLU portrays and how they should be solved, giving insight on what it truly means to translate a text for a non-human reader (a machine), as well as demonstrating the importance of culture in translation. With this study, it was intended to, if not create a definitive style guide for this type of projects, at least pave the way.

That leads to the fourth objective, which was *to propose a course of action for the creation of a final style guide in future editions of the project*. This last objective was completely fulfilled, given that the translation problem analysis aims to provide the project with a set of criteria and guidelines to address the gaps that the original style guide does not cover.

The main objective of this study was *to develop a multilingual style guide that served to translate the MMLU benchmark*. This objective was not fully met for two main reasons. The first reason is that the style guide was created following the instructions of synthetising the Interinstitutional Style Guides (because the project is part of a European initiative and therefore the European Commission's guidelines for translation into the official European languages should be followed), even when they are not aimed at the MMLU's type of text. The initial approach was not to create an ad hoc style guide for this project. The second reason is that initially the intention was to delete every segment of the benchmark that could not be resolved only by following the style guide, that means

deleting every segment with cultural references, which was also not fully achieved. As a result, the creation of the multilingual style guides remains an unmet objective.

5.2. Future Lines of Research

This project has inevitably left some unanswered questions, which could form the basis for future research. Some of these questions are directly applicable to future editions of the project. For instance, the extent to which each type of translation problem affects the evaluation of LLMs, and whether those translation problems that seem "weaker", such as misspellings or typos, truly have no effect on the model's evaluation process. Another aspect worth exploring is whether the thematic heterogeneity of the MMLU hinders the establishment of uniform translation criteria applicable across all segments.

Other questions, however, go beyond the scope of this project and open up broader research questions for the field. A rather philosophical one, with reminiscences of the Ship of Theseus paradox, is whether complete equivalence among multilingual versions of the same benchmark is achievable, even through localisation, or if after making all these changes in a benchmark it can still be said that it is the same one, even when core parts of it (as cultural references) have been changed. Another question that arises from this analysis is to what extent do the translation strategies and techniques developed for human readers apply when translating for machines, and whether new alternatives should be explored.

6. BIBLIOGRAPHY

Abeysinghe, B., & Circi, R. (2024). The challenges of evaluating LLM applications: An analysis of automated, human, and LLM-based approaches. *arXiv preprint arXiv:2406.03339*. https://doi.org/10.48550/arXiv.2406.03339

Anastasiou, D., & Schäler, R. (2010). Translating vital information: Localisation, internationalisation, and globalisation. *Synthèses journal, 3*(11), 11-25. https://www.academia.edu/download/25124449/syntheses.pdf [Accessed: 11/05/2026]

Anthropic. (2026). *Claude* (claude-sonnet-4-6) [Large language model]. https://www.anthropic.com

Bergmann, D. (2025, November 17). What is fine-tuning? IBM. https://www.ibm.com/think/topics/fine-tuning [Accessed: 11/05/2026]

Cao, J., Chan, Y. K., Ling, Z., Wang, W., Li, S., Liu, M., Qiao, R., Han, Y., Wang, C., Yu, B., He, P., Wang, S., Zheng, Z., Lyu, M. R., & Cheung, S. C. (2025). How should we build a benchmark? Revisiting 274 code-related benchmarks for LLMs. *arXiv preprint arXiv:2501.10711*. https://doi.org/10.48550/arXiv.2501.10711

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology, 15*(3), 1-45. https://doi.org/10.1145/3641289

Corte Fernández, N. (2002). Localización e internacionalización de sitios web. *Revista Tradumàtica: Traducció i tecnologies de la informació i la comunicació*, (1). Londres (City University): Electronic Publishing. https://ddd.uab.cat/record/4760

De Beaugrande, R., & Dressler, W. U. (1981). *Introduction to text linguistics*. Longman.

Diego, A. S. (2026, March 11). *Introducción a los LLM, los Grandes Modelos de Lenguaje* (Versión marzo 2026). Adolfo Sanz De Diego. https://www.asanzdiego.com/2026/03/introduccion-los-llm-los-grandes-modelos-de-lenguaje-version-marzo-2026.html [Accessed: 11/05/2026]

European Commission. (n.d.). *Guidelines for translating EU documents into Spanish*. Knowledge Centre on Translation and Interpretation. https://knowledge-centre-translation-interpretation.ec.europa.eu/en/resources-translating-eu-documents/guidelines-translating-eu-documents-spanish [Accessed: 11/05/2026]

European Union. (n.d.a). *Resources for translating EU documents*. Knowledge Centre for Translation and Interpretation. https://knowledge-centre-translation-interpretation.ec.europa.eu/en/resources-translating-eu-documents [Accessed: 11/05/2026]

European Union. (n.d.b). *Style guides*. Eurostat. https://ec.europa.eu/eurostat/web/user-guides/style-guides [Accessed: 11/05/2026]

European Union. (n.d.c). *Official Journal of the European Union: direct access.* EUR-Lex. https://eur-lex.europa.eu/oj/direct-access.html?locale=en [Accessed: 11/05/2026]

European Union. (n.d.d). *Languages*. https://european-union.europa.eu/principles-countries-history/languages_en [Accessed: 11/05/2026]

Farnschläder, T. (2025, January 27). *AI hallucinations: A guide with examples*. DataCamp. https://www.datacamp.com/blog/ai-hallucination [Accessed: 11/05/2026]

Folaron, D. (2025). Translation Technology and Minority Languages in a Multilingual Digital World. In S. Baumgarten & M. Tieber (Eds.), *The Routledge Handbook of Translation Technology and Society*, (pp. 66-85). Routledge. http://dx.doi.org/10.4324/9781003271314-6

Fourrier, C., Frere, T., Penedo, G., & Wolf, T. (2025, December 3). The LLM Evaluation Guidebook. *Hugging face*. https://huggingface.co/spaces/OpenEvals/evaluation-guidebook#the-model-builder-perspective-am-i-building-a-strong-model [Accessed: 11/05/2026]

Ghosal, P. (2026, February 20). *Why you should work with a translation style guide.* Phrase. https://www.phrase.com/blog/posts/translation-why-you-should-work-with-a-style-guide/ [Accessed: 11/05/2026]

Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Al-Garadi, M. A., Hassan, S. Z., Shoma, M., Wu, J., Mirjalili, S., & Shah, M. (2025, February 10). LLMs: A comprehensive survey of applications, challenges, datasets, models, limitations, and future prospects. *TechRxiv*. https://doi.org/10.36227/techrxiv.23589741.v8

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). Measuring massive multitask language understanding. *arXiv preprint arXiv: 2009.03300*. https://doi.org/10.48550/arXiv.2009.03300

Holdsworth, J. (n.d.). *What is AI bias?* IBM. https://www.ibm.com/think/topics/ai-bias [Accessed: 11/05/2026]

Hu, T., & Zhou, X. H. (2024). Unveiling LLM evaluation focused on metrics: Challenges and solutions. *arXiv preprint arXiv:2404.09135.* https://doi.org/10.48550/arXiv.2404.09135

International Organization for Standardization. (2015). *ISO 17100:2015: Translation services — Requirements for translation services*. https://www.iso.org/standard/59149.html

Interpro Translation Solutions. (2024, April 15). C*reating a style guide for translation: A step-by-step guide*. https://www.interproinc.com/creating-a-style-guide-for-translation/#why [Accessed: 11/05/2026]

Jiménez-Crespo, M. A. (2018). Localisation research in Translation Studies: expanding the limits or blurring the lines? In Dam, H., Brogger, M., Zethsen, K. (Eds.), *Moving*

boundaries in translation studies (pp. 26-44). Routledge. https://doi.org/10.4324/9781315121871

Jiménez-Crespo, M. A. (2010). Localization and writing for a new medium: a review of digital style guides. *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, (8), (pp. 1-9). https://raco.cat/index.php/Tradumatica/article/view/225904

Joshi, S. (2025). Evaluation of large language models: Review of metrics, applications, and methodologies. *Preprints*. https://doi.org/10.20944/preprints202504.0369.v1

Karzhev, S. (2024, August 6). *LLM Evaluation: Metrics, Methodologies, Best Practices*. DataCamp. https://www.datacamp.com/blog/llm-evaluation [Accessed: 11/05/2026]

Kumar, R. (2025, June 11). *What is MMLU? LLM Benchmark Explained and Why It Matters.* DataCamp. https://www.datacamp.com/blog/what-is-mmlu [Accessed: 11/05/2026]

Li, H., Dong, Q., Chen, J., Su, H., Zhou, Y., Ai, Q., Ye, Z., & Liu, Y. (2024). LLMs-as-judges: a comprehensive survey on llm-based evaluation methods. *arXiv preprint arXiv:2412.05579*. https://doi.org/10.48550/arXiv.2412.05579

Liu, Y., Yao, Y., Ton, J. F., Zhang, X., Guo, R., Cheng, H., Klochkov, Y., Taufiq, M. F., & Li, H. (2023). Trustworthy llms: a survey and guideline for evaluating large language models' alignment. *arXiv preprint arXiv:2308.05374*. https://doi.org/10.48550/arXiv.2308.05374

Miller, J. K., & Tang, W. (2025). Evaluating LLM metrics through real-world capabilities. *arXiv preprint arXiv:2505.08253*. https://doi.org/10.48550/arXiv.2505.08253

Molina, L., & Hurtado Albir, A. (2002). Translation techniques revisited: A dynamic and functionalist approach. *Meta, 47*(4), 498-512. https://doi.org/10.7202/008033ar

Moyer, U. (2001) Glossaries and Terminology Lists—Their Importance to Quality Translations. In J. Watkins (Ed.), *The guide to translation and localization: Preparing products for the global marketplace* (pp. 16-21). Lingo Systems. https://translationjournal.net/images/e-Books/PDF_Files/The%20Guide%20to%20Translation%20and%20Localization.pdf

Ni, S., Chen, G., Li, S., Chen, X., Li, S., Wang, B., Wang, Q., Wang, X., Zhang, Y., Fan, L., Li, C., Xu, R., Sun, L., & Yang, M. (2025a). A survey on large language model benchmarks. *arXiv preprint arXiv:2508.15361.* https://doi.org/10.48550/arXiv.2508.15361

Ni, S., Kong, X., Li, C., Hu, X., Xu, R., Zhu, J., & Yang, M. (2025b, April). Training on the benchmark is not all you need. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 23, pp. 24948-24956). https://doi.org/10.1609/aaai.v39i23.34678

OpenAI. (2025). *ChatGPT* (March 14 version) [Large Language Model]. https://chat.openai.com

OpenAI: Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., … & Zoph, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*. https://doi.org/10.48550/arXiv.2303.08774

OpenAI: Jaech, A., Kalai, A., Lerer, A., Richardson, A., El-Kishky, A., Low, A., Helyar, A., Madry, A., Beutel, A., Carney, A., Iftimie, A., Karpenko, A., Passos, A. T., Neitz, A., Prokofiev, A., Wei, A., Tam, A., Bennett, A., Kumar, A. … & Li, Z. (2024). Openai o1 system card. *arXiv preprint arXiv:2412.16720*. https://doi.org/10.48550/arXiv.2412.16720

Phrase. (2025). *What does i18n mean? A simple definition of internationalization.* https://phrase.com/blog/posts/i18n-a-simple-definition/ [Accessed: 11/05/2026]

Publications Office of the European Union. (2022a). *Interinstitutional style guide*. Publications Office of the European Union. https://data.europa.eu/doi/10.2830/215072 [Accessed: 11/05/2026]

Publications Office of the European Union. (2022b). *Libro de estilo interinstitucional*. Publications Office of the European Union. https://data.europa.eu/doi/10.2830/31113 [Accessed: 11/05/2026]

Publications Office of the European Union. (n.d.a). *Interinstitutional Style Guide: General publications*. https://style-guide.europa.eu/en/content/-/isg/topic?identifier=part-two-general-publications [Accessed: 11/05/2026]

Publications Office of the European Union. (n.d.b). *Part Three: Conventions common to all languages.* Interinstitutional Style Guide. https://style-guide.europa.eu/en/content/-/isg/topic?identifier=part-three-conventions-common-to-all-languages [Accessed: 11/05/2026]

Publications Office of the European Union. (n.d.c). *Part Four: Publications in a specific language. Interinstitutional* Style Guide. https://style-guide.europa.eu/en/content/-/isg/topic?identifier=part-four-publications-in-a-specific-language [Accessed: 11/05/2026]

Schäffner, C. (1998). Skopos theory. In M. Baker (Ed.), *Routledge encyclopedia of translation studies*, (pp. 235-238). Routledge.

Singh, S., Romanou, A., Fourrier, C., Adelani, D. I., Ngui, J. G., Vila-Suero, D., Limkonchotiwat, P., Marchisio, K., Leong, W. Q., Susanto, Y., Ng, R., Longpre, S., Ruder, S., Ko, W.-Y., Bosselut, A., Oh, A., Martins, A., Choshen, L., Ippolito, D., Ferrante, E., Fadaee, M., Ermis, B., & Hooker, S. (2025, July). Global MMLU: Understanding and addressing cultural and linguistic biases in multilingual evaluation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 18761-18799). https://doi.org/10.18653/v1/2025.acl-long.919

Stryker, C. (2025, November 17). *What are large language models (LLMs)?* Ibm.com. https://www.ibm.com/think/topics/large-language-models [Accessed: 11/05/2026]

TÜV SÜD America, Inc. (n.d.). *ISO 17100 & ISO 18587 – Translation quality standards*. https://www.tuvsud.com/en-us/services/auditing-and-system-certification/iso-17100 [Accessed: 11/05/2026]

Vingert, T. (2021, April 19). *6 tips to write an effective translation brief*. Women in Localization. https://womeninlocalization.com/6-tips-to-write-an-effective-translation-brief/ [Accessed: 11/05/2026]

Zhuang, Z., Chen, Q., Ma, L., Li, M., Han, Y., Qian, Y., Bai, H., Zhang, W., & Liu, T. (2023, August). Through the lens of core competency: Survey on evaluation of large language models. In J. Zhang (Ed.), *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 2: Frontier Forum)*, (pp. 88-109). Chinese Information Processing Society of China. https://aclanthology.org/2023.ccl-2.8/ [Accessed: 11/05/2026]

7. APPENDIX

STYLEGUIDES

V.01, 16 February 2026

MMLU Translation project EMT – DGT AI Unit

Sabela Calvet Alvarez

Coordinated by WG5 – Susana Valdez and Pilar Sánchez Gijón

- This style guide aims to help translators in their journey translating the MMLU evaluation dataset.
- Here you will find a list of language-related items to which special attention must be paid while translating, as well as the links to the Interinstitutional EU Guidelines where you will be able to find relevant information regarding said topics.
- In no case can AI or NMT be used to translate the content of the dataset.
- Languages described in the following table are those involved in the MMLU dataset first translation project (February-June 2026).
- This styleguide includes the following sections:

1. Country designations and abbreviations
2. Languages
3. Rules for expressing monetary units
4. Punctuation
5. Word-processor punctuation marks and spacing
6. Capitalisation (upper and lower case)
7. Numbers, dates and time
8. Abbreviations
9. Acronyms and/or initialisms
10. Italics
11. Verbs
12. Spelling
13. Prepositions
14. Use of articles
15. Miscellaneous

	BG	ES	CS	DE	ET	EL	EN	FR	GA	IT	LV	HU	NL	PL	PT	SL
1. Country designations and abbreviations	7.1.1. Наименования и съкращения, които следва да бъдат използвани	7.1.1. Denominaciones y siglas preceptivas	7.1.1 Názvy a zkratky	7.1.1. Zu verwendende Länderbezeichnungen und Kürzel	7.1.1. Kasutusel olevad nimetused ja lühendid	7.1.1. Ονομασίες και συντομογραφίες	7.1.1. Designations and abbreviations to use	7.1.1. Dénominations et sigles à utiliser	7.1.1. Ainmneacha agus giorrúcháin a úsáidtear	7.1.1. Denominazioni e sigle da utilizzare	7.1.1. Izmantojamie apzīmējumi un abreviatūras	7.1.1. Megnevezések és rövidítések	7.1.1. Benamingen te gebruiken afkortingen	7.1.1. Stosowane nazwy i skróty	7.1.1. Denominações e siglas a utilizar	7.1.1 Imena in kratice
2. Languages	7.2.1. Последователност на езиковите версии и ISO кодове (много езични текстове)	7.2.1. Orden de las versiones lingüísticas y códigos ISO (textos multilingües)	7.2.1 Kódy ISO a pořadí jazykových verzí (vícejazyčné texty)	7.2.1. Reihenfolge der verschiedenen Sprachfassungen eines Textes und ISO-Codes (mehrsprachige Texte)	7.2.1. ISO koodide ja keeleversioonide järjestus mitmekeelsetes tekstides	7.2.1. Σειρά γλωσσών και κωδικοί ISO (πολύγλωσσα κείμενα)	7.2.1. Order of language versions and ISO codes (multilingual texts)	7.2.1. Ordre des versions linguistiques et codes ISO (textes plurilingues)	7.2.1. Ord leagan acha teanga agus cód ISO (Eagraíocht Idirnáisiúnta gCaighdeán) (téacsanna ilteangacha)	7.2.1. Ordine delle versioni linguistiche e codici ISO (testi multilingui)	7.2.1. Valodu versiju secība un ISO kodi (vairākvalodu teksti)	7.2.1. Nyelvek és ISO-kódok sorrendje (többnyelvű szövegrészek)	7.2.1. Volgorde van de taalversies en ISO-codes (meertalige teksten)	7.2.1. Kolejność wersji językowych wyste pujących w tym samym tekście oraz kody ISO (teksty wielojęzyczne)	7.2.1. Ordem das versões linguísticas e códigos ISO (documentos plurilingues)	7.2.1 Oznake ISO in vrstni red jezikovnih različic v večjezičnih besedilih
3. Rules for expressing monetary units	7.3.3. Правила за изписване на	7.3.3. Normas de escritura de	7.3.3 Pravidla pro označování	7.3.3.Schreibregeln für Währu	7.3.3. Rahaühikute kasutamise	7.3.3. Κανόνες γραφής	7.3.3. Rules for expressing	7.3.3. Règles d'écriture des référe	7.3.3. Rialacha maidirle	7.3.3. Regole di scrittura dei	7.3.3. Noteikumi par naudas	7.3.3. A pénznemek írásán	7.3.3. Schrijfwijze van munte	7.3.3. Zasady zapisu jednos	7.3.3. Regras de escrita das	7.3.3 Pravila o izraža nju

	BG	ES	CS	DE	ET	EL	EN	FR	GA	IT	LV	HU	NL	PL	PT	SL
	парични единици	las referencias monetarias	měnových jednotek	ngsbezeichnungen	eeskirjad	νομισματικών μονάδων	monetary units	nces monétaires	haonaid airgeadaíochta a chur in iúl	riferimenti monetari	vienību attēlošanu	ak szabályai	enheden	tek monetarnych	referências monetárias	denarnih enot
4. Punctuation	10.1 Пунктуационни правила	10.1 Puntuación	10.1 Interpunkce	10.6. Klammern 10.10. Strich	10.1 Kirjavahemärgid	10.1 Στίξη	10.1 Punctuation	10.1 Ponctuation	10.1. An phonc aíocht	10.1 Punteggiatura	10.1 Interpunkcija	10.1 Írásjelek	10.1 Leestekens	10.1 Interpunkcja	10.4 Pontuação	10.1 Ločila
5. Word-processor punctuation marks and spacing	6.4. Разстояние между знаците при текстообработка	6.4. Espaciado de los signos de puntuación	6.4 Interpunkce a mezery ve Wordu	6.4. Abstände bei der Zeichensetzung	6.4. Wordi tekstitöötlusprogrammi kirjavahemärgid ja sõnavahe	6.4. Διαστήματα και σημεία στίξης	6.4. Word-processor punctuation marks and spacing	6.4. Espacements des signes de ponctuation	6.4. Marcanna poncaíochta próiseálaithe focal, agus spásáil	6.4. Spaziatura dei segni d'interpunzione	6.4. Interpunkcijas zīmes un atstarpes datoru apstrādātekstā	6.4. Központozás és tagolás szövegszerkesztő használata esetén	6.4. Spatie s bij interpunctie	6.4. Znaki interpunkcyjne i spacje w edytorze Word	6.4. Espaços dos sinais de pontuação	6.4 Ločila in presledki pri urejevalnikih besedil
6. Capitalisation (upper and lower case)	10.2. Главни букви	10.2. Mayúsculas y minúsculas	10.4 Psaní velkých písmen	No specific information. Follow general rules in German	10.3. Suur ja väike algustäht	10.3. Κεφαλαία και μικρά	10.4. Upper and lower case	10.2. Majuscules et minuscules	10.4. Ceannlitreacha agus cás íochtair	10.2. Maiuscole e minuscole	10.4. Lielo sākumburtu lietojums	10.2. Kis és nagy kezdőbetűk	10.2. Hoofdletters	10.2. Użycie wielkiej i małej litery	10.7. Maiúsculas e minúsculas	10.4 Velika in mala začetnica
7. Numbers, dates and time	No specific information.	10.3. Números	10.5 Čísla, data a čas	10.3. Datum	10.4. Numbrid, kuupäe	10.8. Αριθμοί, ημερο	10.5. Numbers, dates	10.4. Nombres	10.3. Uimhreacha, dátaí	10.3. Numeri	10.5. Skaitļi, datumi	10.3. Rövidítések és	10.5. Getallen	10.8. Liczebniki	10.9. Nomes numer	10.5 Datum in čas

	BG	ES	CS	DE	ET	EL	EN	FR	GA	IT	LV	HU	NL	PL	PT	SL
	Follow general rules in Bulgarian				evad ja kellaaeg	μηνίες και ώρες	and time		agus an t-am		un laiks	mozaikszók 10.6. Számok és keltezés			ais (números)	
8. Abbreviations	10.4.2. Съкращения	10.4. Abreviaturas	10.7 Zkratky, zkratková slova, symboly a akronyma	10.1. Abkürzungen	10.6. Lühendamine	10.6. Συντομογραφίες	10.8. Abbreviations and symbols	10.5. Abréviations	10.5. Giorrúcháin, coimrithe úcháin, acrainmneacha agus siombailí	10.4. Abbreviazioni	10.7. Saīsinājumi, simboli un akronīmi	10.4. Idegennevek, szavakírása és toldalékolása	10.4. Afkortingen	10.3.1. Skróty	10.10. Abreviaturas	10.7 Kratice
9. Acronyms and/or initialisms	10.4.1. Абревиатури	10.5. Siglas y acrónimos	10.7 Zkratky, zkratková slova, symboly a akronyma	10.2. Kürzel und Akronyme	10.8. Lühendid ja akronüümid	10.7. Αρκτικόλεξα	10.8. Abbreviations and symbols	10.6. Sigles et acronymes	10.5. Giorrúcháin, coimrithe úcháin, acrainmneacha agus siombailí	10.5. Sigle e acronimi	10.7. Saīsinājumi, simboli un akronīmi	10.5. Egybeírás és különírás	No specific information. Follow general rules in Dutch	10.3.2. Skrótowce (akronimy) 10.3.3. Zgoda między podmiotem a orzeczeniem 10.3.4.	10.11. Siglas e acrónimos	10.7 Kratice

	BG	ES	CS	DE	ET	EL	EN	FR	GA	IT	LV	HU	NL	PL	PT	SL
														Określanie rodzaju skrótowców na podstawie ich wymowy		
10. Italics	—	—	10.6 Kurziva	—	10.5. Kaldkiri	10.2. Χρήση πλάγιων χαρακτήρων	10.7. Italics	—	10.6. Cló iodálach	—	10.6. Slīpraksts	—	—	—	—	10.6 Ležeča pisava
11. Verbs	—	—	10.2 Slovesa	—	—	—	10.2. Verbs	—	—	—	—	—	—	—	—	10.2 Glagol
12. Spelling	10.3. Слято, полуслято и разделно писане на сложните и съставните думи	—	—	—	—	—	10.3. Spelling	—	10.2. An litriú	—	10.2. Vārdu un vārdformu izvēle 10.3. Pareizrakstība	—	10.3. Los, aaneen of streepje	10.4. Zestawienia – z łącznikiem czy bez 10.5. Razem czy osobno	10.5. Prefixos	—

	BG	ES	CS	DE	ET	EL	EN	FR	GA	IT	LV	HU	NL	PL	PT	SL
13. Prepositions	10.5.5. Предлози	—	—	—	—	—	—	—	10.8. Réamhfhocail agus frásaí réamhfhocla cha	—	—	—	—	10.9. Przyimki	10.6. Preposições	—
14. Use of articles	10.5.1. Членуване	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15. Miscellaneous	—	—	10.3 Pravopis, stylistika a mluvnice	10.5. Gradzeichen 10.8. Prozent- und Promillezeichen	10.7. Mitmesugust	10.9. Προβλήματα από όσης στην ελληνική γλώσσα 10.4. Το τελικό νι	—	—	10.9. Nithe ilchineálacha	—	10.8. Dažādi	10.7. Egyéb	—	10.10. Zaimek „swój” 10.12. Fleksja 10.13. Inne wyrażenia sprawiające problemy	10.8. Particularidades da lingua gem	10.8 Razno

Web Resources

- Guidelines for EU translation into Bulgarian
- Guidelines for EU translation into Czech
- Guidelines for EU translation into Dutch
- Guidelines for EU translation into English
- Guidelines for EU translation into Estonian
- Guidelines for EU translation into French
- Guidelines for EU translation into German
- Guidelines for EU translation into Greek
- Guidelines for EU translation into Hungarian
- Guidelines for EU translation into Irish
- Guidelines for EU translation into Italian
- Guidelines for EU translation into Latvian
- Guidelines for EU translation into Polish
- Guidelines for EU translation into Portuguese
- Guidelines for EU translation into Slovenian
- Guidelines for EU translation into Spanish