

European e-Justice Portal Translation Quality Assessment

JUST/2024/PR/JACC/CRIM/0141

Specific Contract No. 141

under Framework Contract COMM/2020/OP/0024 – Lot 2

Authors: StreamLine consortium

Translation Quality of the e-Justice Portal

Comparative Assessment of Human and Machine Translation

Executive Summary

The assessment of the European e-Justice Portal's translation quality, conducted across all 23 official EU languages, compared human and machine translation (MT) using blind evaluation against five criteria: accuracy, terminology, linguistic correctness, consistency, and readability.

Key Findings:

- **Human Translation:**
Human translations consistently achieve high standards, with an overall average score of 4.2/5. Strengths included linguistic correctness (4.5/5), consistency (4.4/5), and readability (4.2/5). Minor weaknesses are noted in terminology (4.0/5) and accuracy (4.0/5), mainly due to occasional inconsistencies or overly literal translations. Human translators reliably preserve legal meaning and intent, with only minor stylistic or terminological errors.
- **Machine Translation (MT):**
MT output is generally comprehensible and linguistically correct (4.1/5), but lag behind as regards accuracy (3.5/5), terminology (3.3/5), and consistency (3.8/5), with an overall average score of 3.6/5. Persistent issues include mistranslations, inconsistent terminology, and literal or mechanical phrasing.
- **Comparative Assessment:**
Human translation outperforms MT across all criteria, especially in terminology, consistency, and readability. MT is approaching near-human performance for simple, standard texts, but human expertise remains essential for legal accuracy, nuanced language, and context-sensitive translation.

Introduction

The European e-Justice Portal offers multilingual content for various EU audiences. High-quality translations are crucial for legal reliability and accessibility. This assessment is part of a broader evaluation and usability study with two main goals:

- to understand user engagement and the Portal's practical usefulness (Impact Assessment), and
- to review content structure, language accessibility, usability, and translation quality, recommending improvements (Quality Review).

The translation quality review specifically aims to determine whether the translations on the e-justice Portal are sufficiently accurate, consistent, and legally reliable across all 23 official EU languages, while remaining accessible and understandable to diverse audiences.

The translation quality review also aims to compare the performance of human and machine translation. To achieve this, selected pages translated by human translators were translated using machine translation and subsequently submitted to a blind assessment.

The assessment was conducted across all language pairs of the e-justice Portal and included a combination of quantitative scoring and qualitative expert comments. Reviewers evaluated each translation against five main quality criteria:

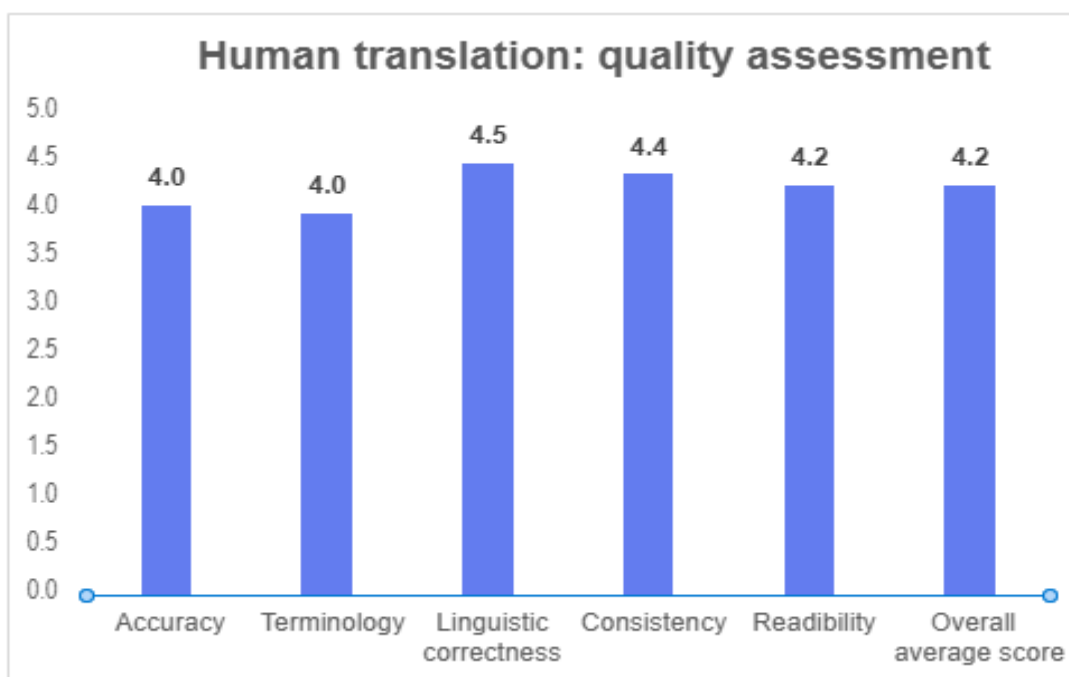
- accuracy
- terminology
- linguistic correctness
- consistency
- readability

The evaluation was conducted blindly: reviewers were not informed whether a text had been produced by a professional human translator or by a machine translation system. This approach ensured that judgments were based solely on the quality of the translations, trying to eliminate potential bias toward or against machine-generated output.

1. Human Translation Quality Assessment

The analysis demonstrates that human translation achieves a high standard of professional quality. Human translations show precision, linguistic fluency and strong understanding of the context. While minor inconsistencies and some issues in terminology may occasionally occur, the overall results confirm the reliability and readability of human-produced translations.

1.1. Human translation: Overall performance



The overall performance of human translation is strong, with an average total score of 4.2, reflecting a high standard of quality. Linguistic correctness stands out as the top strength, achieving the highest rating at 4.5 and demonstrating excellent adherence to grammar and language rules. Consistency (4.4) and readability (4.2) are also rated high, indicating that translations are both coherent and easy to understand.

However, terminology receives the lowest score at 4.0, suggesting some challenges with the use of specialised legal vocabulary or term accuracy. Similarly, accuracy is rated at 4.0, indicating the need to improve the precision in conveying meaning.

1.2. Performance by criterion

1.2.1 Accuracy

Human translators generally provide accurate and reliable legal translations. Most of the time, the translated texts keep the original meaning, logic, and structure. Reviewers often noted that translators correctly understood and expressed legal procedures and concepts.

Some issues appear regularly:

- Small parts of the text—like short clauses or extra notes—are left out or added.
- There are occasional mistakes in translating the names of institutions or official references (for example, “Provinzgericht” or “Qorti tal-Magistrati”).

- Translators sometimes simplify or interpret the text to make it clearer, which can slightly reduce technical accuracy.

Despite these minor issues, no serious mistakes that altered the meaning of the texts were spotted. In fact, reviewers noted that when translators made changes, these usually helped make the information clearer and easier to understand.

Human translators reliably preserve legal intent and factual integrity. Accuracy errors are rare and typically stylistic or terminological, not conceptual.

1.2.2. Terminology

Terminology presents certain challenges, with an average score of 4.0 out of 5. The strongest results for this criterion were observed in language pairs such as French-Spanish, French-Portuguese, Portuguese-Spanish, Spanish-Portuguese, Italian-Bulgarian and Greek-Bulgarian, where translators demonstrate a solid grasp of the legal vocabularies of both source and target languages. However, for other language pairs we can note a variable level of quality as regards terminology and occasional inconsistencies.

Positive aspects:

- Most translators showed a strong understanding of legal concepts and generally used terminology consistent with national legal practice.
- Several translations, particularly between Spanish and Portuguese (both ways), demonstrate excellent consistency and discipline in the use of legal terms.

Key issues:

- Inconsistent use of terms within the same document (e.g., “reclamaciones” vs. “demandas”; “izdevumi” translated as “costs,” “expenses,” or “charges”).
- Mixing of terminology from different legal traditions, such as British and continental European systems.
- Occasional lack of research into jurisdiction-specific terminology or omission of official terms.
- Translations sometimes too literal, use of a word-for-word translation that may not be familiar or appropriate.

1.2.3. Linguistic Correctness

Among all criteria, linguistic correctness demonstrated the strongest performance, with an average score of 4.5 out of 5 and many near-perfect evaluations (5 out of 5).

Human translators displayed very good grammatical control, syntax, and orthographic precision. Common observations include:

- Flawless punctuation and capitalisation, particularly in French, Italian, and Spanish translations.
- Grammatically complete and balanced sentences.
- Appropriate use of tenses, with coherent sentence structures.

A few issues were identified, including minor punctuation inconsistencies, mixed apostrophes, or overly complex syntax inherited from the source language. However, these issues are rare and generally minor, affecting only the style rather than the overall quality of the translation.

1.2.4. Consistency

Consistency achieved an average score of approximately 4.4 out of 5, indicating that the majority of translations maintain a high level of consistency. In most cases, translators apply the same terminology, formatting, and structural patterns throughout the text.

Some translations—especially those in Bulgarian, Spanish, and German—showed almost perfect consistency, with a uniform legal tone, formatting, and style throughout the selected text.

Key issues:

- Differences in how words are capitalised or abbreviated (for example, “Art.” versus “Article”).
- Slight changes in tone between sections, such as switching from a formal to a more casual style.
- Formatting issues, like missing spaces, inconsistent numbering, or uneven spacing in lists.

1.2.5. Readability

Readability, with an average score of **4.2 out of 5**, reflects the translators’ ability to render complex legal material into natural, idiomatic, and accessible prose.

Positive aspects:

- Sentences flow naturally, respecting the structure and logic of the target language.
- Legal terminology maintained while remaining comprehensible to non-specialist audiences.
- Style often adapted effectively to audiences.

Key issues:

- Some translations described as **“heavy” or “bureaucratic”**, showing excessive literalness or source interference.
- Overly complex sentence structures in certain languages (notably German and Polish).
- A few cases of **repetition or redundancy**.

1.3. Conclusions

The average score of human translation is around 4.2 out of 5. All key dimensions meet or exceed the threshold of acceptability for publication, requiring minor editorial improvement.

Key Insights

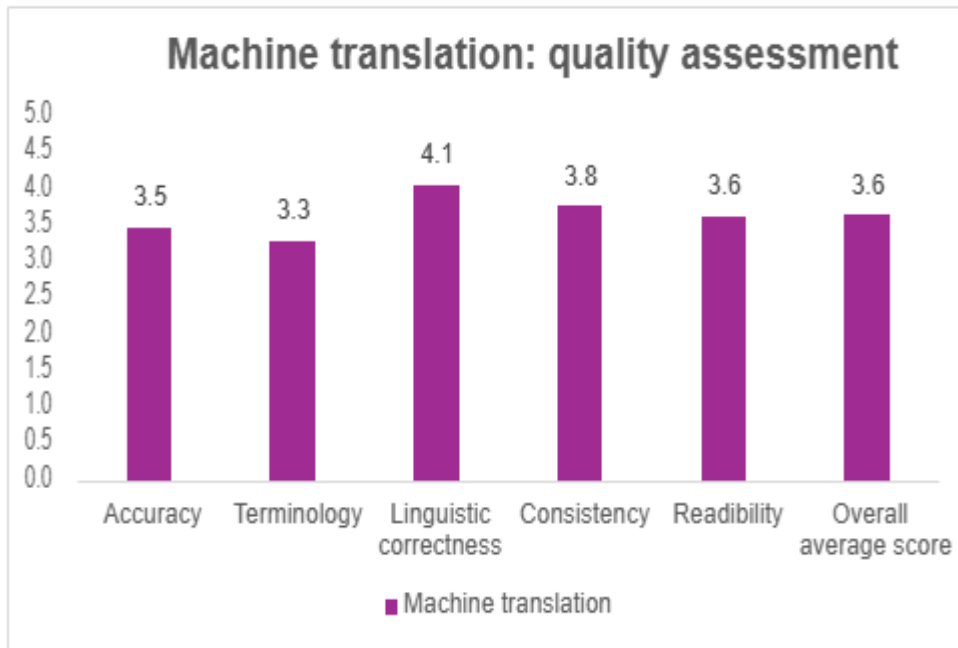
- Human translation ensures faithful, precise communication of legal concepts.
- Terminology and consistency remain the most frequent points of improvement.
- Readability, while high, could be further improved by promoting plain legal language strategies.

Human translation demonstrates the professional standard expected for multilingual legal texts: accurate, grammatically correct, context-aware, and stylistically coherent.

2. Machine Translation Quality Assessment

The evaluation indicates that machine translation (MT) of legal texts generally produces grammatically and linguistically sound output that is broadly comprehensible. Nonetheless, persistent issues with accuracy and terminology continue to pose significant challenges.

2.1. Machine translation – Overall performance



Based on the evaluation results, the machine translation of legal texts demonstrates generally strong linguistic correctness and moderate readability, with respective scores of 4.1 and 3.6 out of 5. The translations are consistent (3.8/5) and reasonably accurate (3.5/5). However, the use of specialised terminology remains a significant challenge, as reflected in a lower score of 3.3 out of 5. Overall, the average total score of 3.6 indicates that, while machine translation can produce comprehensible and mostly correct legal texts, persistent issues with accuracy and terminology warrant further improvement.

2.2. Performance by criterion

2.2.1. Accuracy

Accuracy represents the most critical challenge for MT in legal translation. Reviewers repeatedly highlight **mistranslations, omissions, additions, and conceptual distortions**, including:

- Incorrect rendering of modality (“may” translated as “must”), fundamentally altering legal obligations.
- Misinterpretation of procedural terminology and legal entities (e.g., *Statio bail*, *Garda Siochána*, *bankruptcy decree*).
- Additional text not present in the source text compromising factual fidelity. Ex: “where the value of a claim does not exceed EUR 5,000”.

Recurrent mistakes:

- Literal word-for-word that distorts intended meaning.
- Insertion or omission of phrases affecting the meaning of legal norms.
- Lack of alignment between source and target sentence structures leading to inconsistencies.

Accuracy remains to be a critical concern for machine translation of legal texts. Although MT reproduces most factual elements correctly, the risk of misinterpretation of obligations or rights may cause issues for users.

2.2.2. Terminology

Terminology received the **lowest average score (3.3 out of 5)** and attracted the highest concentration of critical comments, making it the weakest aspect of the translations. Legal texts rely heavily on precise, jurisdiction-specific vocabulary; deviations may alter meaning or render the translation non-compliant with national practice.

Key issues:

- **Inconsistent translation of core legal terms**, e.g., “payment order” vs. “order for payment”, *izdevumi* translated variously as “costs,” “expenses,” and “expenditures.”
- **Non-standard equivalents** due to lack of domain adaptation (e.g., *responsabilité pénale* inconsistently rendered, *ouderlijk gezag* mistranslated).
- **Overly general language**, weakening legal precision (“procedure” used generically instead of specific procedural terms).
- **Inadequate handling of multilingual EU legal systems**, where terms differ across jurisdictions (e.g., Austrian vs. German terminology).

While some translations, notably Polish and Bulgarian ones, demonstrated partial domain awareness and proper use of national legal terminology, overall MT terminology remains **unstable and unreliable** without post-editing.

Terminological inconsistency undermines MT’s use for professional legal translations. Post-editing should focus on terminology and legal references to achieve high-quality legal translation.

2.2.3. Linguistic Correctness

Linguistic correctness—grammar, syntax, and spelling—scored the highest (4.1 out of 5), indicating substantial progress in MT’s ability to produce linguistically sound content. Reviewers often note:

- Correct punctuation, capitalisation, and orthography.

- Grammatically correct structures, though sometimes overlong or awkward.
- Rare but noticeable errors (e.g., verb tense mismatches, missing articles).

Reviewers also observe that syntactic correctness does not necessarily imply fluency. Many sentences are technically correct but mechanical, with rigid structure and unnatural rhythm due to literal translation of the source language content.

While machine translation is generally sufficient for basic understanding, linguistic correctness alone does not guarantee quality. Translated texts often sound unnatural or non-native, especially in longer or more complex legal passages.

2.2.4. Consistency

Consistency, with an average of 3.7 out of 5, reflects a relatively low level of performance in this dimension.

Key issues:

- Inconsistent rendering of repeated terms (*payment order* vs. *order for payment*).
- Variations in formatting, including capitalisation, punctuation, and abbreviation (“Art.” vs. “Article”).
- Unstable tone, oscillating between formal and informal expressions.

Consistency issues are particularly visible in longer content, where alignment of legal terminology across sections is essential.

Consistency is adequate but insufficient for professional standards. While MT systems generally maintain structural coherence within short passages, they do not retain **memory** of previously used terminology and stylistic choices, leading to avoidable variations.

2.2.5. Readability

The readability criterion receives a relatively low score of 3.6 out of 5. Reviewers consistently describe MT texts as literal, rigid, or mechanical. Sentences tend to follow the source language structure too closely, leading to unnatural sentence and occasionally confusing syntax.

Key issues:

- “Too literal in places, which affects readability.”
- “Long and convoluted sentences make the text difficult to read.”
- “Very poor flow, not readable or comprehensible for the Dutch user.”

Even when technically accurate, the tone of the translated text does not match the tone used in the source text: it is sometimes too colloquial, sometimes overly bureaucratic. However, several reviewers note that content remains understandable, even if stylistically inelegant.

MT output is appropriate for general understanding. However, stylistic post-editing is required to ensure the text sounds idiomatic and suitable for both professional and general audiences.

2.3. Conclusions

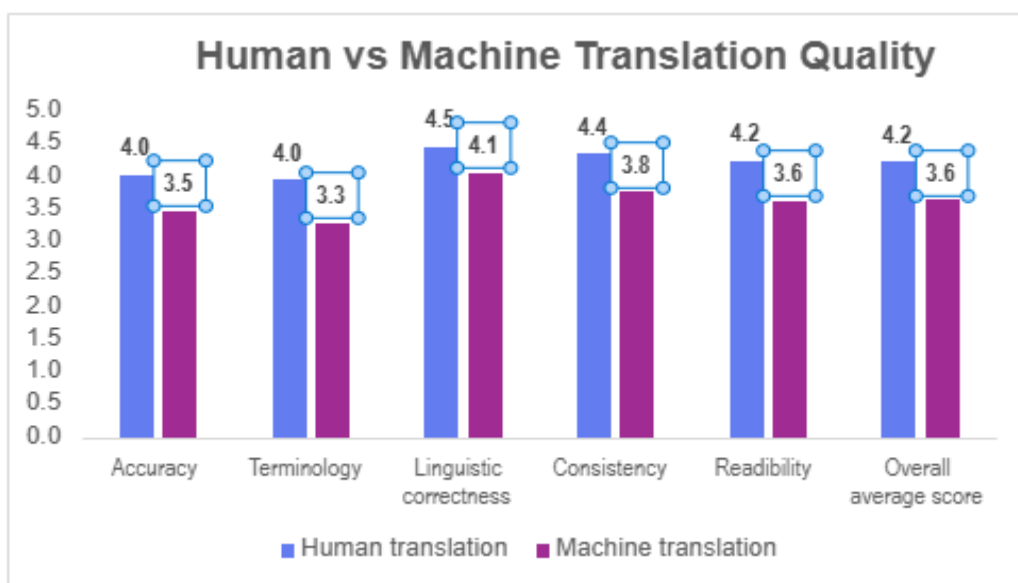
The evaluation indicates that the quality of machine translation for legal texts is improving, with linguistically sound and generally understandable content output. However, accuracy and terminology weaknesses remain major obstacles to professional use without human revision. Addressing these areas will be key to achieving fully reliable legal translations.

3. Human translation versus machine translation

3.1. Methodology

The evaluation of both human and machine translations was conducted as a blind assessment, meaning that reviewers were not informed about the origin of each translation. This ensured that assessments were based exclusively on the intrinsic quality of the translations, with the view to eliminate any preconceptions about human or machine performance.

3.2. Human vs Machine Translation: overall performance



Human translation outperforms machine translation across all evaluated criteria, particularly as regards linguistic correctness, consistency, and readability. While machine translation achieves an acceptable level of linguistic correctness (4.1) and accuracy (3.5), it lags behind as regards terminology (3.3) and overall quality (3.6 compared to 4.2 for human translation). This implies that, although machine translation can produce generally understandable and structurally sound content, human translation remains superior, especially for texts requiring high precision, nuanced language, and consistent terminology.

3.2.1. Human translation still outperformed machine translation

Despite not knowing which texts are human or machine generated, reviewers consistently assigned higher scores to human translations across all categories. Human translations averaged approximately 4.2 out of 5, compared with 3.6 out of 5 for machine translations.

This confirmed that human translators remain superior in producing accurate, contextually consistent, and stylistically appropriate output.

3.2.2. Linguistic correctness

Both human and machine translations achieved high scores for linguistic correctness:

- Human translation: 4.5
- Machine translation: 4.0

This indicates that MT models have now reached a high level of quality in grammar, punctuation, and sentence structure.

The main difference between human and machine translation lays in terminology, semantic nuances, and natural tone, where human expertise remains superior.

3.2.3. Some MT outputs achieved near-human ratings

In several language pairs (e.g., EN–ES, EN–DE), certain MT outputs received scores similar to human translations, particularly for linguistic correctness and readability. This demonstrates that for standardised and structurally simple texts, MT could approach human-level fluency.

However, as soon as the text required legal reasoning, contextual adaptation, or subtle lexical choices, the advantage shifted decisively back to human translators.

3.3. Conclusions

The blind assessment confirmed that machine translation performed well in linguistic correctness and superficial readability, producing outcomes that seem qualitative on first

reading. However, it frequently displayed terminological inconsistency, literal translation, and contextual imprecision that undermined the quality of the translation.

In contrast, human translators continue to demonstrate a deeper understanding of semantic nuance, greater control over legal and procedural terminology, and superior ability to produce idiomatic, natural, and contextually appropriate language.

Recommendations

- Continue to rely on professional human translators for legal and procedural texts, especially those requiring nuances, precise terminology, and contextual adaptation.
- Use machine translation for initial drafts or simple content, complemented by human post-editing to ensure the use of appropriate and precise terminology, legal accuracy and high stylistic quality.
- Ensure the systematic use by translators of up-to-date multilingual terminology databases (IATE, [Juremy](#)) providing translation for specific legal terms to reduce inconsistencies.
- Encourage translators to use clear, accessible language to improve readability for non-specialist audiences, while maintaining legal precision when it is feasible and does not alter the legal content.
- Put in place user feedback collection and disseminate the feedback to DGT. It may serve to take corrective actions.
- Encourage authors of original text to apply EC editorial guidelines ([Europa Web Guide](#) and the “[How to Write Clearly](#)”) when drafting the original version. This will contribute to drafting clearer and more readable content, which will be an important factor in ensuring high-quality translations.