

FRAMEWORK FOR TRANSLATION QUALITY METRICS¹

LETRINT PROJECT

(“Legal Translation in International Institutional Settings”)²

Overview

This framework classifies 14 **error types** into three major categories (Section 1) based on [Prieto Ramos \(2015\)](#)³ and a comprehensive review of academic and professional approaches to translation quality, which included the Multidimensional Quality Metrics ([MQM](#)), international translation standards and quality assurance policies applied in several institutions.⁴

Each identified error is assigned a degree of impact and a **penalty score** according to the scale described in Section 2. Finally, average penalty scores (penalty points per 1000 words of source text) are correlated to **quality levels** from a holistic perspective in line with the expectations of professional translations in international institutional settings in particular (Section 3).

While this framework was primarily developed for the LETRINT project, it is also compatible with the approaches reviewed and was designed to be applicable for other translation assessment purposes.⁵

¹ How to cite this document: Prieto Ramos, Fernando (2023). *Framework for Translation Quality Metrics – LETRINT Project*. Available from: <https://transius.unige.ch/en/research/letrint/quality-metrics>.

² Led by Fernando Prieto Ramos (PI), University of Geneva, and supported by the Swiss National Science Foundation through a Consolidator Grant (157797).

³ Prieto Ramos, Fernando (2015). “Quality Assurance in Legal Translation: Evaluating Process, Competence and Product in the Pursuit of Adequacy.” *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, 28(1), 11–30. <https://doi.org/10.1007/s11196-014-9390-9>.

⁴ The project PI would like to thank all institutional informants for their valuable cooperation in sharing their expertise and relevant materials, in particular, in alphabetical order, John Beaven (Council of the EU), Cecilia Elizalde and Anne Lafeber (United Nations), Hannah Riley (European Economic and Social Committee and European Committee of the Regions), Ingemar Strandvik (European Commission), Angelika Vaasa (European Parliament) and Susan Wright (Court of Justice of the EU). Special thanks go to Diego Guzmán for his continued support in the process of testing and refining the framework.

⁵ See also Prieto Ramos, Fernando & Guzmán, Diego (2024). “The impact of specialised translator training and professional experience on legal translation quality assurance: an empirical study of revision performance.” *The Interpreter and Translator Trainer*, 18(2), 313–337. <https://doi.org/10.1080/1750399X.2024.2344948>.

1. ERROR CATEGORIES & TYPES⁶

(1) **ACCURACY** (content reformulation errors associated with mistranslations or incomplete translations, including **content** completeness)

1.1. Unjustified omission

1.2. Unjustified addition

1.3. Opposite meaning

1.4. Incorrect data: figure, date, reference (e.g. piece of legislation or provision number), proper name (e.g. body, person, programme, acronym) or other factual information⁷

1.5. Other incorrect meaning or inaccuracy (not included above)

(2) **LINGUISTIC CORRECTNESS / LANGUAGE MECHANICS AND STYLE** (errors that occur due to deviations from **language** norms; focus is on language dynamics and form rather than semantic impact)

2.1. Grammar or cohesion issue (e.g. number and gender agreement, word repetition)

2.2. Spelling

2.3. Punctuation or typographical error

2.4. Register or idiomaticity issue

2.5. Other style, fluency or readability issues not included above (e.g. inappropriate / awkward word collocation or sentence length)

(3) **TERMINOLOGICAL AND FUNCTIONAL ADEQUACY** (errors arising from inadequate conformity to established **terminology**, applicable domain or institutional **conventions**, unjustified intratextual inconsistencies or other inappropriate translation decisions from a **macrotextual functional perspective**)

3.1. Incorrect or inappropriate terminology (i.e. when not conforming to specifications or to established terminology that is relevant and expected in a particular domain or setting)⁸

3.2. Non-conformity to phraseological or genre convention (e.g. established expressions or formulations that are characteristic of a particular genre or required according to applicable legal instruments, institutional guidelines or models)

3.3. Unjustified terminological or phraseological inconsistency (i.e. multiple renderings for the same concept or expression where consistency should be preserved throughout a text)

3.4. Other adequacy issues, such as inadequate reformulation or translation procedures in light of the purpose of the target text (e.g. inadequate use of a culture-specific reference, a borrowing or a descriptive translation, or a rendering of a segment for which a previous translation should be used (e.g. cases of quotation); the reformulation may be accurate and linguistically correct, but inappropriate for the purpose of the translation from a functional perspective)

⁶ No design, markup or format issues are evaluated for the purpose of the project.

⁷ Usually minor formal corrections of errors associated with details overlooked in the translation process.

⁸ While terminology issues can be classified as accuracy or language errors (or treated as a separate error category as is the case in other assessment models), here they are considered a matter of adequacy for the particular purpose of the translation, considering the relevant communicative needs, as well as the applicable job specifications, domain conventions or institutional standards.

2. SEVERITY LEVELS

Each identified error is assigned a degree of impact and a penalty score according to the scale below. The **impact and penalty** of an error type may vary depending on the context and the text purpose, e.g. a punctuation error may have a critical semantic impact in a particularly sensitive segment or text, but no significant impact in other cases. No error is considered “neutral” in this framework. However, **repeated errors** are not assigned penalty points, e.g. a repeated accuracy or terminological issue is only penalized once in the same text. Punctuation errors, e.g. missing commas, are not considered repeated errors and are therefore penalized in each instance, unless they appear in identical segments in a text.

- **Minor (-1): no significant impact on content, text function or overall comprehension or fluency**, i.e. the error is obvious and does not affect these aspects. For example: “Nada de lo dispuesto en el presente Convenio menoscabarán [...]” instead of “menoscabará [...]” (our emphasis) for “Nothing in this Convention shall affect [...]” (wrong subject-verb agreement).
- **Significant (-2): limited impact on content or fluency**, i.e. the error causes a change of semantic nuance or hinders readability but the affected element does not significantly impact the overall scope or function of the text as a whole considering the relevance and context of the segment. For example: “1º de septiembre de 2005” instead of “1º de septiembre de 2004” (error in the first of several references to the effective date of a legal change).
- **Major (-3): serious impact on content, comprehension or text function**, i.e. the error may lead to a change in scope or understandability of important elements of the text, such as key concepts, definitions, conditions or deadlines. For example: “carezca de ánimo de lucro” instead of “tenga ánimo de lucro” for “organised on a for-profit basis” (opposite meaning in the definition of entities that are subject to a legal act).
- **Critical (-5 / DISQ.): very serious or disqualifying errors** that may render the text unfit for use or even involve legal, political or economic risks. For example: “870,94 millones de dólares por concepto de intereses” instead of “870 941 dólares por concepto de intereses” for “\$870,941 in interest” (incorrect expression of an amount to be reimbursed).

This taxonomy, as well as the penalty values, can be adapted depending on assessment purposes and can be reduced, for example, to three categories: minor, significant and major.

3. OVERALL QUALITY LEVELS

The penalty scores (penalty points per 1000 words of source text) are correlated to **quality levels** from a holistic perspective (based on Prieto Ramos, 2015: 23–26). As in the case of severity levels, the average impact suggested here can be adapted to varying evaluation purposes or quality expectations. For example, a poor translation of level 2 or D might be usable for basic informal needs, but unfit for institutional purposes, while a borderline translation of level 3 or C may correspond to a pass or fail value in a training context depending on the course level and text difficulty, and the same translation may be considered inadequate for professional use. These quality levels can accommodate variations tailored to the aims of each assessment, e.g. by adjusting the penalty scores and/or assessment results associated to each level.

Excellent (A/5) – average impact of 0-1 penalty points per 1000 words

High accuracy and consistency, adequate decisions according to the legal conditions and communicative situation, no mistranslation or linguistic error.

Acceptable (B/4) – average impact of 1-3 penalty points per 1000 words

Only few minor inaccuracy, inconsistency, inadequate decision or linguistic errors not affecting main text functions or microtextual communicative priorities.

Borderline (C/3) – average impact of 4-6 penalty points per 1000 words

Inadequate translation decision or significant linguistic error or several minor ones (e.g. punctuation problems) hinder comprehension or microtextual priorities but do not affect the main text functions; the translation may be usable depending on its purpose.

Poor (D/2) – average impact of 7-9 penalty points per 1000 words

One major problem of accuracy, adequacy or language usage, or two significant errors, and/or multiple minor ones, have a serious impact on content, comprehension or text functions, rendering the text unfit for publication, even if it may be readable.

Unacceptable (E/1) – average impact of 10 penalty points or more per 1000 words

Critical error, inaccurate content, systematically inadequate decision-making or numerous linguistic errors render the text unfit for any use.

Quality level	Penalty score*	Error types and severity	Impact of errors	Target text usability
Excellent (A/5)	0-1	None or a marginal (minor) error	None	Fit for publication
Acceptable (B/4)	1-3	Few minor errors (-1)	Minor	Fit for publication if needed
Borderline (C/3)	4-6	One significant error (-2) AND/OR Several minor errors (-1)	Comprehension or microtextual priorities affected, but not main text functions	Usable depending on purpose; publication not recommended without revision
Poor (D/2)	7-9	At least one major error (-3) or two significant errors (-2) AND/OR Multiple minor errors (-1)	Content, comprehension or text functions affected significantly	Unfit for publication even if text may be readable
Unacceptable (E/1)	10 ≤ or DISQ.	One critical error of high impact (-5 & DISQ.) OR Numerous errors of various severity levels, including at least one major error (-3) or two significant errors (-2)	Content, comprehension or text functions seriously compromised	Unfit for any use

* Penalty points per 1000 words of source text.