MARCH 2026

EU Court of Justice Orders NMT Engine: Pilot Project Proposal

PRESENTED TO:

Court of Justice of the European Union
Language Services Localization Manager

PRESENTED BY:

Araomis Medina Alejandro, Zaiana Banzaraktsaeva, Katerin Torres Gomez, Codi McBride

Objectives

The purpose of this study is to assess the feasibility of employing a custom-trained NMT engine to translate Court of Justice Orders from the European Union from English into Spanish. The engine will provide translations that preserve meaning, maintain procedural validity, allow for human translation without the risk of legal misinterpretation and meet the following criteria:

Timing

PEMT 50% faster than Human Translation



Pricing

PEMT 25% cheaper than Human Translation



Quality

PEMT will be evaluated using the LISA QA Metric on a sample of 2,000 words per evaluation cycle.

To pass:

- No Critical errors
- A maximum weighted score of 25 error points per 1,000 words.
- No more than 10 total points from Mistranslation errors per 1,000 words.

Scores exceeding these thresholds will be considered a failure. Results from each evaluation cycle will inform further training and tuning of the MT engine during the pilot phase.

LISA QUALITY METRIC

Error type	Minor	Major	Critical
Mistranslation	1	5	10
Accuracy	1	5	10
Terminology	1	5	10
Language	1	5	10
Style	1	5	10
Country	1	5	10
Consistency	1	5	10

Source:
http://producthelp.sdl.com/SDL_TMS_2011/en/Creating_and_Maintaining_Organizations/Managing_QA_Models/LISA_QA_Model.htm

Timeline

1. MARCH 1 — Deliver Pilot Proposal
2. MARCH 4 — Client Meeting
3. MARCH 5 - APRIL 2 — NMT training
4. APRIL 3 - APRIL 4 — Data Analysis
5. APRIL 5 — Deliver Updated Proposal

Training and Data

10 Rounds

Testing Data
Orders of the Court of Justice 2025-2026

Tuning data
Orders of the Court of Justice 2023 - 2024

Training Data
Orders of the Court of Justice 1955 - 2022

Workflow

1. **Data Acquisition:** Collect sufficient data, 5,000 words, on the Orders of the Court of Justice of the European Union in the EUR-Lex corpora.

2. **Cleaning and Alignment:** Align segments and paragraphs in both languages (ENG-SPA) with SDL Trados, remove duplicates, and standardize numbers, citations, and name entities with Olifant.

3. **Baseline Training:** Start training with legal document data in English and Spanish using Microsoft Custom Translator.

4. **Domain Tuning:** Tune only Orders for 2023-24, enter domain-specific prompts with Microsoft Custom Translator to improve our BLUE score.

5. **Testing:** Add Order documents from recent years, 2025-2026.

6. **Human Evaluation:** Verify whether the engine exceeds our pass/fail score to identify what we need to change in order to have a successful MT engine that meets our requirements.

Costs

Tasks	Estimated Hours	Rate	Cost
Data Collection	6	$40	$240
Segmentation, Alignment, Cleaning	20	$40	$800
MT Training Rounds	20	$40	$800
Post-Editing	8	$40	$320
QA	4	$40	$160
Human Evaluation	6	$40	$240
TOTAL			**$2560**

Deliverables

Updated Proposal for training NMT engine for EU Court of Justice orders translation which will include:

- Details of data used in training
- BLEU score evaluations of each round
- Final cost of the project
- Evaluation of overall feasibility of project and time/cost
- Suggestions for further training of the NMT engine