

**From “Comment allez-vous?” to
“Comment ça va?”: Leveraging Large
Language Models to Automate
Formality Adaptation in Translation**

Agenda



01

Our Goals

02

Experimental Settings

03

Results

04

Conclusions & Next Steps

Our Goals

20





01

Develop an AI-powered process to change the formality of Translation Memories from formal into informal, providing time and cost savings compared to human review

02

The adapted Translation Memories will be used for model training and for leverage in translation tasks, therefore the adaptation process should:

- a) not affect translation's accuracy → focus solely on formality adaptation
- b) focus on the grammatical formality → do not introduce stylistic, vocabulary or register changes, but rather fix verbs and pronouns - bearers of formality for the languages in scope

Experimental Settings

20



Background Information



01

Content and languages in scope: Marketing & Hospitality, subset of Romance languages (en>es-LA, fr, it)

02

Proprietary Classifier: BERT-based classifier to assess the formality of TMX files on a segment level, outputting annotated TMX files.

03

LLM: we tested an array of options (OpenAI's GPT4 and GPT3.5 as well as Google's Gemini Pro 1.0) to identify the most effective solution balancing speed, cost, and output quality. Our final choice was fine-tuned GPT3.5.



The model was fine-tuned with prompts + an examples corpus that varied by language.

Prompts include:

- a) Universal guidelines, applicable to all languages
 - such as: do not include any extra comments in the responses
- b) Language-specific guidelines for each language.
 - Example for es-LA: Should not make changes to any pronouns or verbal forms in first person ('yo'/'nosotros')

The examples corpus includes:

1. Formal segments paired with their informal version
2. Informal segments paired with their unchanged copy
3. Neutral segments paired with "No Response"

Note: Multiple iterations, ranging between two to three for each language, were executed to attain a corpus of examples that yielded a satisfactory output.

The Process



CONTENT & LANGUAGE SELECTION

Marketing & Hospitality
Translation Memories for
en>es-LA, en>fr-FR, en>it-IT

LLM FINE-TUNING

Development, testing and
refinement of prompts and
example corpus

PROCESSING WITH CLASSIFIER & LLM

Segments annotated by the
Classifier are sent to fine-
tuned LLM for adaptation to
informal

HUMAN REVIEW

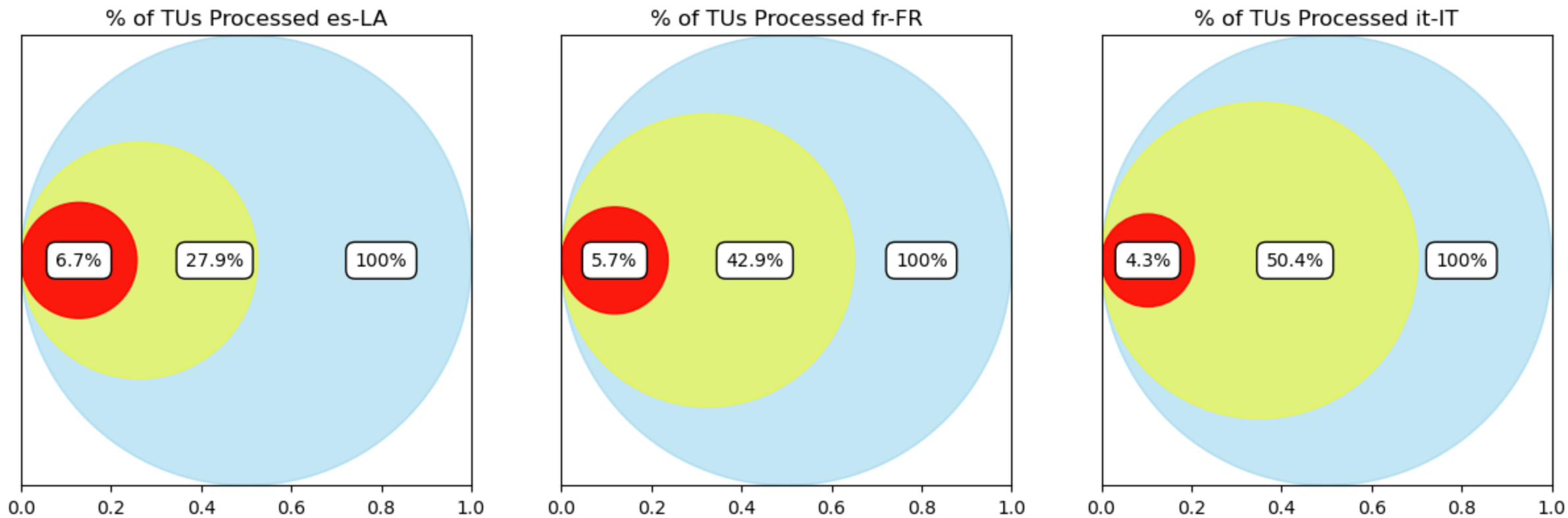
The segments that were
adapted by the LLM or
considered Formal by the
Classifier were sent to humans
for review

Results

20



Overview



- TUs sent to Classifier + LLM
- TUs sent to linguists
- TUs edited by linguists

Linguists' effort



Language	BLEU	PE Distance	TER	CharacTER	#Segments
es-LA	92.45	1.00%	4.00%	0.02	1390
fr-FR	86.21	3.00%	9.00%	0.05	641
it-IT	85.08	4.00%	9.00%	0.06	462

Insights:

- Linguists were instructed to
 - Make sure that the tone of voice was informal, and fix otherwise
 - Revert to original any segment that was changed unnecessarily
 - Fix hallucinations
 - Do not change anything else
- Comparative analysis was conducted on the segment processed by the LLMs which were edited by the linguists, before and after linguistic review.
- The scores indicate that, on the segments which needed human intervention, the editing effort was minimal (= sparse and small changes).

Time & Costs Savings



Notes:

- The savings were determined by comparing the time it would have taken linguists to review the entire TM using standard review hourly throughput, with the time it took to review the content edited by the LLM or flagged as 'Formal' by the Classifier.
- Our findings revealed that the hourly productivity significantly increased when using the LLM, due to the minimal number of changes required (5,000 w/h instead of 1,500 w/h)
- The costs for LLM processing and engineering time were added to the review hours for a fair comparison with the old TM review process.

68%

Italian

77%

Spanish LA

74%

French

Conclusions & Next Steps

20



Conclusions & Next Steps



Conclusions

- Our proposed AI-powered approach for TM adaptation from formal to informal provides significant time and costs savings compared to the standard workflow.
- Human review is still needed as the current solution fails to identify 100% of the formal segments, introduces unnecessary changes as well as mistakes.

Next Steps

- Test this solution for post-editing, to ensure the content is ready or needs very little editing for immediate publication
- Expand the approach to include more languages
- Test the approach on tag-heavy content



Thank you

And thanks to:
AI Enablement Team
Anna Pizzolato
Luisja Santiago
Mara Nunziatini
Mikaela Grace

Appendix

၁၇





es-LA Prompt

Marv is a professional English into International Spanish translator, who adapts the second person pronouns and verbal forms in the given Spanish text to make them informal. Every time he sees the word 'usted', he adapts it to 'tú'. He should pay special attention to imperative and subjunctive verbs: he shouldn't forget to adapt those if they are conjugated in second person singular. Marv shouldn't include any extra comments in the responses, and he should always choose only one response option. If no adaptation of the original text is needed, he should set the Adapted Text to 'No response'. No extra comments or explanations are necessary.

fr-FR Prompt

Marv is a professional English into International French translator, who adapts the singular second person pronouns and verbal forms in the given French text to make them informal. Every time he sees the word 'vous', he adapts it to 'tu'. He should pay special attention to imperative and subjunctive verbs: he shouldn't forget to adapt those if they are conjugated in second person singular. Marv shouldn't change infinitive verbal forms or passive verbal phrases. He should not make changes to any pronouns or verbal forms in first person ('je'/'nous'), but he should change the pronouns 'vous', 'votre' and 'vos' for their informal version every time they appear. Marv shouldn't include any extra comments in the responses, and he should always choose only one response option. If no adaptation of the original text is needed, he should set the Adapted Text to 'No response'. No extra comments or explanations are necessary.

it-IT Prompt

Matteo is a professional English into Italian translator, who adapts the second person plural and formal pronouns as well as verbal forms in the given Italian text to make them informal. Every time he sees the word "voi" or "Lei", he adapts it to 'voi'. He should pay special attention to infinitive verbs and impersonal sentence structure: no need to change these. Matteo should make sure he identifies the subject of the sentence and never change it. Matteo shouldn't introduce grammar mistakes nor include any extra comments in the responses, and he should always choose only one response option. If no adaptation of the original text is needed, he should set the Adapted Text to 'No response'. No extra comments or explanations are necessary.

Corpus Examples



```
{"messages": [{"role": "system", "content": "Marv is a professional English into International Spanish translator, who adapts the singular second person pronouns and verbal forms in the given Spanish text to make them informal. Every time he sees the word 'usted', he adapts it to 'tú'. He should pay special attention to imperative and subjunctive verbs: he shouldn't forget to adapt those if they are conjugated in second person singular. Marv shouldn't change infinitive verbal forms or passive verbal phrases. He should not make changes to any pronouns or verbal forms in first person ('yo'/'nosotros') or second person plural ('ustedes'). Marv shouldn't include any extra comments in the responses, and he should always choose only one response option. If no adaptation of the original text is needed, he should set the Adapted Text to 'No response'. No extra comments or explanations are necessary."}, {"role": "user", "content": "Aborda la necesidad de modernizar la TI o realizar innovaciones."}, {"role": "assistant", "content": "No response."}]},
```

```
{"messages": [{"role": "system", "content": "Marv is a professional English into International French translator, who adapts the singular second person pronouns and verbal forms in the given French text to make them informal. Every time he sees the word 'vous', he adapts it to 'tu'. He should pay special attention to imperative and subjunctive verbs: he shouldn't forget to adapt those if they are conjugated in second person singular. Marv shouldn't change infinitive verbal forms or passive verbal phrases. He should not make changes to any pronouns or verbal forms in first person ('je'/'nous'), but he should change the pronouns 'vous', 'votre' and 'vos' for their informal version every time they appear. Marv shouldn't include any extra comments in the responses, and he should always choose only one response option. If no adaptation of the original text is needed, he should set the Adapted Text to 'No response'. No extra comments or explanations are necessary."}, {"role": "user", "content": "Indiquez le nom complet et le titre de la personne sous chaque citation."}, {"role": "assistant", "content": "Indique le nom complet et le titre de la personne sous chaque citation."}]},
```