

# GenAI in Localization

João Graça





# João Graça

CTO & Co-founder, **Unbabel**

Member of the **Forbes Technology Council**

Founder, **Lisbon Machine Learning School**

PhD in Natural Language Processing

University of Pennsylvania & Instituto Superior Técnico



# João Graça

Happily married

Proud father of 3 amazing princesses

Enthusiastic surfer and padel player

Is the localization industry  
going to disappear?

# Google Translate Boasts Near-Human Accuracy Levels With New Neural Network-Powered System

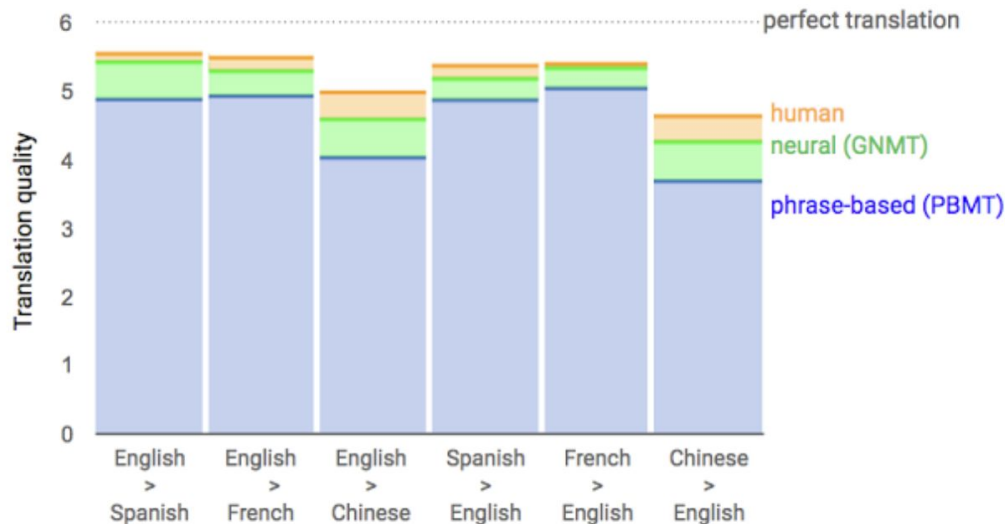
News

By [Lucian Armasu](#) published September 27, 2016



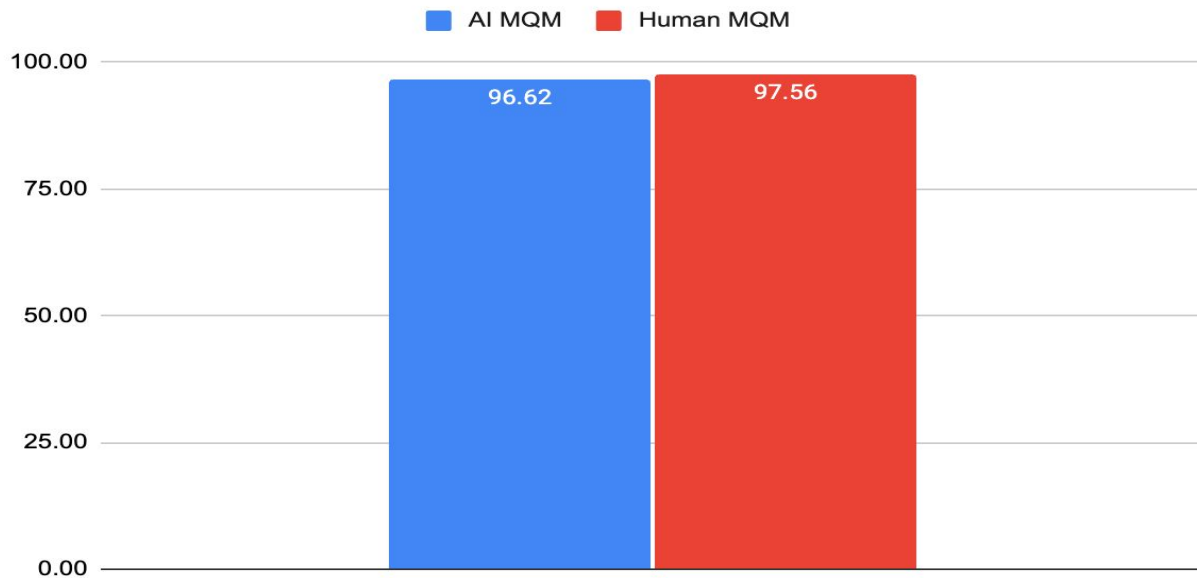
Comments (9)

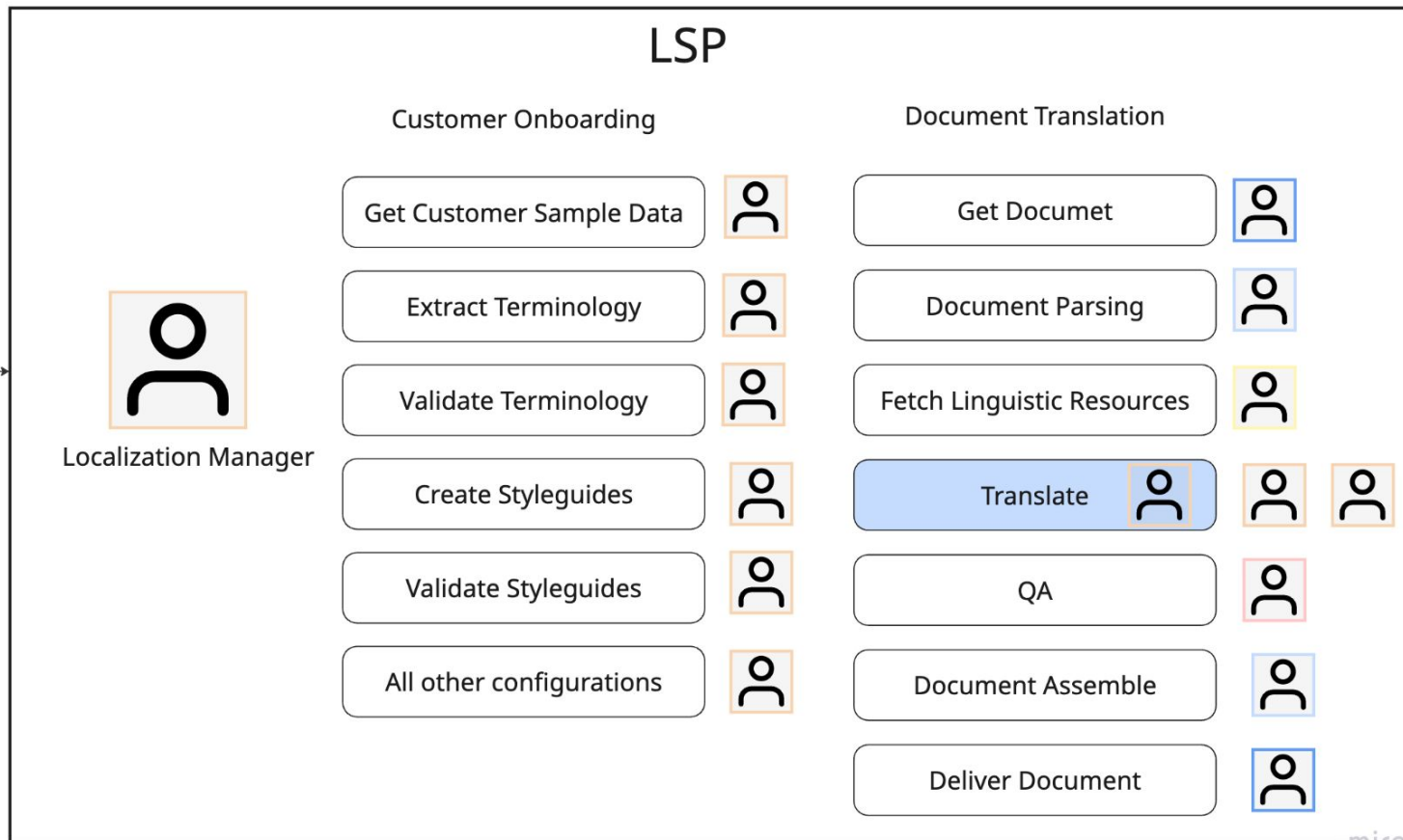
When you purchase through links on our site, we may earn an affiliate commission. [Here's how it works.](#)



# Startup CEO says humans won't be needed for translation in 3 years as it launches AI app

PUBLISHED WED, NOV 13 2024•12:56 AM EST | UPDATED WED, NOV 13 2024•1:57 AM EST









# Machine Translation Goes Agentic

TransAgents, a system that boosts literary translation with a multi-agent workflow

Machine Learning Research

Generative AI

Large Language Models (LLMs)

Agents

📅 Published

Aug 14, 2024

🕒 Reading time

3 min read



“The human work involved in localization  
will change but not disappear”



# What's my role in future?

PM, Translator, QA, FE....

# AI Translation

# Humans vs Machine

## **AI**

**Unlimited use of AI  
tools**

**No human  
involvement**

**Money is no object**

**Custom Prompts**

## **Humans**

**Use Unbabel's  
resources**

**Linguists**

**Editor instructions  
and selection**

# Text Samples

## Literary

*“In the morning when thou risest unwillingly, let this thought be present—I am rising to the work of a human being”*

## Technical

*“A loss of voltage or a complete power outage can range from downed power lines to failed transformers, to tripped circuit breakers, etc. These power outages may last for a short one-half cycle of the 60 hertz sine wave (8.3 milliseconds) or for several hours.”*

## Legal

*”17-2-0101 Generally. The “R”, residential districts are intended to create, maintain and promote a variety of housing opportunities for individual households and to maintain the desired physical character of the city's existing neighborhoods.*

## Translation Guidelines:

The goal is to translate old English version of "Meditations" into modern, accessible {target\_language} while retaining the philosophical essence and integrity of Marcus Aurelius's original thoughts.

### 1. **Tone and Register**

- Maintain a reflective, personal tone: Meditations is a deeply personal text, with Marcus Aurelius often speaking to himself as a way to contemplate life, ethics, and leadership. The translator should preserve this intimate, contemplative tone.
- Avoid overly formal or archaic language: Although it is an ancient text, using overly formal or outdated terms can alienate modern readers. The language should feel timeless, not antiquated.

### 2. **Philosophical Terminology**

- Translate Stoic concepts with precision: Stoicism is central to Meditations, and key Stoic terms (e.g., logos, virtue, fate, reason, nature) need to be carefully translated. Make sure that the philosophical meaning is preserved and aligned with current scholarly interpretations.
- Use consistent terms: Ensure consistency with how important Stoic concepts like apatheia (freedom from passion) or eudaimonia (flourishing) are translated throughout the text.

### 3. **Clarity and Simplicity**

- Aim for clarity without losing the depth: While Meditations deals with complex philosophical ideas, the text itself often uses simple, direct language. Avoid overly complex or academic translations that might obscure the practical wisdom.
- Simplify where necessary: Marcus Aurelius' language can sometimes be dense or fragmented. Translators should feel free to clarify difficult passages, ensuring they are understandable to modern readers without distorting the original meaning.

# Challenge



**3 types of text**  
(legal, literary and technical).

**2,500 words long.**

**Best scored LPs.**

**Identical instructions.**

**Two weeks to setup**  
translation flows and  
submit.

**Annotators**  
**reviewed with**  
**MQM.**

**Blind evaluations.**



## MQM-Extension : Context-Aware Annotation Framework

Contextual Triggers (source)	MT Errors (target)
Discourse Marker	Agreement
Ellipsis	Collocation
Greeting	Lexical Selection
Lexical Ambiguity	Named Entity
Multiword Expression	Overly Literal
Named Entity	Register
Reference	
Register	
Terminology	

Table 1: **Nine** categories to be applied on the source side, **Contextual Trigger**; **Six** categories to be applied on the **Contextual MT error**.

We might not need humans to translate but we need humans to build the systems that translate

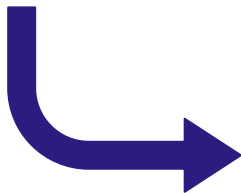
- Build the right prompt
- Set up proper evaluation cycle
- Define finer grained typologies
- Annotate data

# Transcreation

## This response might seem **correct**

**Akane, Japanese Customer**

I'm sorry but the book I ordered isn't here yet!  
What's happening?



**Jamie, Customer Support**

Hi Akane 😊

I'm sorry to hear that your book hasn't arrived yet and that you can't contact the seller. Send me your order number and email address so I can investigate this for you, and please don't cancel your order with them before talking to me first.

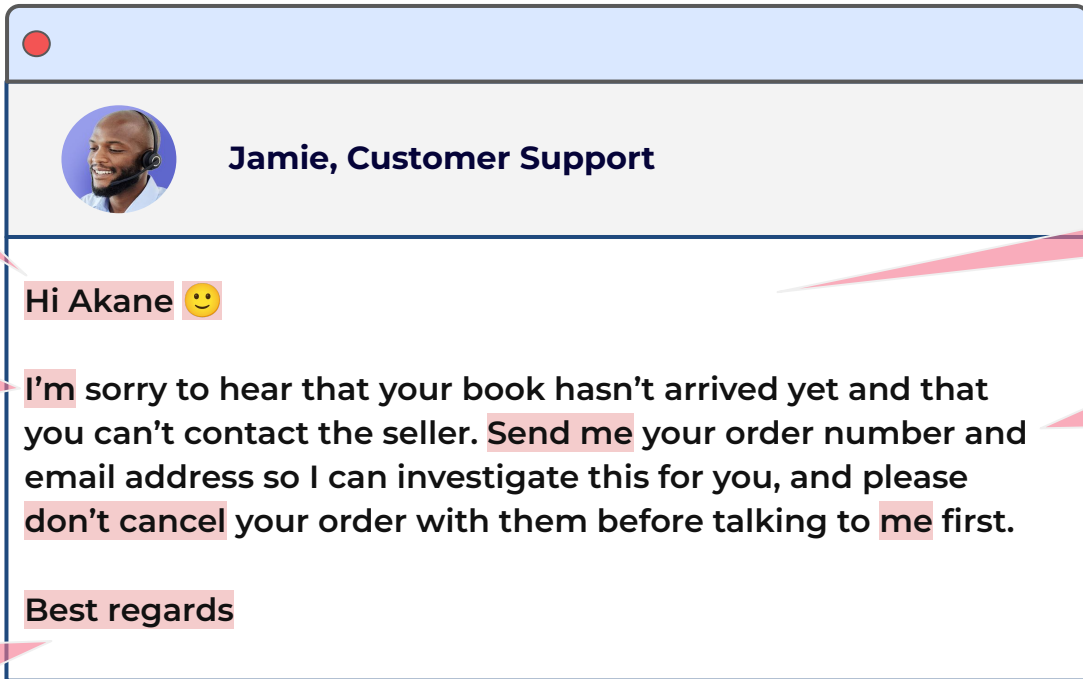
Best regards

## But it can be inappropriate

Should avoid using emojis

Should avoid using the first person singular ("I") - use the first person plural instead ("We")

Should have proper closing remarks.



Requires proper greetings structure

Should avoid giving orders to the reader.

A nighttime photograph of a Japanese street scene. On the left, a building features a large red sign for 'ドラッグ' (Drug) and a smaller sign for 'ハードラッグ' (Hard Drug). A canal runs along the street, reflecting the vibrant neon lights. On the right, more buildings are visible with various signs, including one that says '回ってまっせ!' (Turn around!). The scene is illuminated by a mix of warm and cool colors from the city lights.

Dear Ms. Akane,

Thank you for contacting us.

We apologize for the inconvenience caused by the delay in the delivery of your book. We understand that you have been unable to contact the seller and kindly ask you to provide us with your order number and email address. We would like to ask you to refrain from cancelling the order before speaking to us.

We appreciate your understanding and cooperation.



# Greetings/Closings, Register and Useful Expressions

## Greetings and Closings (Formal)

### Greetings

Emails should start with a greeting that addresses the customer politely and avoids colloquial language. In the case of *chat*, and depending on the domain, the formality of the greeting should not be as high.

When sending *emails* to Japanese customers, instead of greeting the customer with Hello [name], a better way to open an email is by using the customer's name followed by the honorific 様 and directly thanking them in the line below. In English, this should look like in the example below:

❌ Hi John,  
 ✔️ Mr. John,  
 Hello! [email]

If the name of the customer is not known, they can be addressed as "お客様" (Esteemed Customer).

### Register and Honorifics

When addressing customers in Japanese it's important to use the right **register** of formal language and the proper **honorifics**.

As stated in the [Unbabel Language Guidelines](#), Japanese has three levels of formality: *teineigo*, *sonkeigo* and *kenjougo*.

While *teineigo* is a formal register, it represents a more conventional and casual tone and does not place the speaker in a superior or subordinate position. *Kenjougo* and *sonkeigo* are used to, respectively, humble yourself (the speaker) and honor the addressee and convey that the speaker is in a subordinate position.

As such, when talking to customers the *sonkeigo* register should be used to address and talk about the addressee, and the *kenjougo* register should be used to talk about yourself or people in your circle. More detailed information about the use of each type of register can be found in the [Unbabel Language Guidelines](#).

In addition to register, it's important to use the right honorifics to address customers. The appropriate way to address a customer is either by using お客様 ("Valued/Esteemed customer") or the customer's name followed by 様 (example: John様 ["Mr. John"]).

### Closings

Depending on the context, it's appropriate to close emails by thanking the customer for the time they dedicated to the interaction or for their cooperation. In other cases, it may be better to show availability to keep assisting the customer. The following sentences are good examples of email closings:

- 何卒よろしくお願い申し上げます (よろしくお願い致します)。  
I appreciate your kind cooperation
- ご迷惑おかけしますが、よろしくお願い致します。  
I am sorry for your inconvenience, but thank you for your cooperation.
- 他にもご不明な点がございましたら、お気軽にお問い合わせください。  
If you have any (other) questions, please feel free to contact us.
- 貴重なお時間をありがとうございました / ございます。  
Thank you for giving us your valuable time.

### Other useful expressions

- ○○○と申します。  
I am/This is ○○○.
- ご不便をおかけして申し訳ございません。  
I apologize for the inconvenience.
- 申し訳ございませんが  
I apologize but.../We are very sorry but...
- お手数ですが  
恐れ入りますが  
We are sorry for the inconvenience but...
- あいにく  
Unfortunately...
- この度はお問い合わせ頂きありがとうございます。  
Thank you for your inquiry.
- ご返信ありがとうございます。  
Thank you for your response.
- (このたびは) お問い合わせいただき、ありがとうございます。  
Thank you for sharing your concern.
- お客様に不快な思いをさせてしまいましたことをお詫び申し上げます。  
I apologize for your unpleasant experience you went through with us.
- 少しでもお役に立ちましたら幸いです。  
I hope this will help.

# New Typology for Evaluation





The same skill set  
applies to different  
tasks

- Build the right prompt
- Set up proper evaluation cycle
- Define finer grained typologies
- Annotate data

# Terminology Extraction

Pretend you are a professional terminologist.

I want to create a terminology base for the company Unbabel.

For that I wanted you to do the following tasks:

- 1 - Look for public assets of Unbabel, webpages, social media, faq.
- 2 - Extract terms that you believe are specific to the brand Unbabel to form their terminology
- 3 - Find their translation on the Unbabel site public translations

Show the results in a table that is easily copied to excel with the following rows for each term: Term, Definition, Transalctions



Unbabel Terminology Base

Term	Definition	Translation (French)
Language Operations (LangOps)	A holistic function that helps global businesses effectively communicate with their multilingual clients and stakeholders. LangOps provides transparency and control over every aspect of language across a business, utilizing language analytics, augmented machine translation tools, collaborative workflows, seamless integrations, and APIs.	Opérations linguistiques (LangOps)
Human-in-the-Loop AI	An AI-powered language function refined by human translators to create a quality assurance layer in translations, reducing bias and grounding machines in reality.	IA avec intervention humaine
Language Converter	A multilingual, neural translation tool that translates text from one language to another, allowing customer service agents to communicate with customers in different languages.	Convertisseur de langue
Live Chat Translation	Software that automatically translates verbal or written text in real-time to simulate human conversation, enhancing customer service interactions.	Traduction de chat en direct
Localization Services	Services that help organizations adapt their content to local audiences, dialects, or cultural nuances, ensuring the product or service aligns with the target market's expectations.	Services de localisation

Certainly! Based on Skyscanner's publicly available resources, I've compiled a terminology base featuring key terms specific to Skyscanner's brand and services. Each term includes a definition and its translation in French, as found on Skyscanner's French website.

You can easily copy this table into Excel for further use.

Skyscanner Terminology Base

Term	Definition	Translation (French)
Direct Booking Platform	The section of the Skyscanner site where users can book travel and related options directly with the provider, previously known as the 'Facilitated Booking Platform'. ( <a href="#">Skyscanner Partner Terms</a> )	Plateforme de réservation directe
TravelPro	Skyscanner's corporate travel search platform provided to customers for facilitating the booking of travel services. ( <a href="#">Skyscanner Corporate Travel Services Terms</a> )	TravelPro
Result Set	The data set provided to the end user by Skyscanner in response to a search request, including pricing and hyperlinks. ( <a href="#">Skyscanner Partner Terms</a> )	Ensemble de résultats
Price Alerts	A feature that allows users to receive notifications about price changes for specific flights or routes. ( <a href="#">Skyscanner Guide</a> )	Alertes de prix
Everywhere Search	A tool that helps users find the cheapest flights to any destination by searching 'everywhere'. ( <a href="#">Skyscanner Guide</a> )	Recherche "Partout"

If you need translations in other languages or additional terms, feel free to ask!

Learn how to play with  
different models.

Prototype and iterate

- Have you GenAI workspace env.
- Spend time playing with the models
- Move from the simple toy examples like this one to real use cases by following the recipe
  - Evaluation Framework
  - Gather Data
  - Experimentation Loop

# Project Management



Localization Manager

# Language Operations Platform

## Customer Onboarding

Get Customer Sample Data



Extract Terminology



Validate Terminology



Create Styleguides



Validate Styleguides



All other configurations



Prompt management



## Document Translation

Get Documet



Document Parsing



Fetch Linguistic Resources



Translate



QA



Document Assemble



Deliver Document



miro



## How to project manage in the age of agents?

- Learn a couple of no-code agentic frameworks
- Start replacing old flows with agentic flows
- Build missing agents
- Be the coach

# Summary

## What should I do in the next 5 years

- AI is a solid choice for most tasks in the localization industry.
- AI needs to be set up properly
  - Prompting
  - Evaluation
  - Data
- Agentic workflows have a lot of potential but still in experimentation phase. Someone needs to shepherd the agents

What should I do in the  
next 5 years

- You hard skills are still very important
- AI is being deployed across industries
  - Your skills are fundamental for a successful deployment
- Be the LangOps for the rest of the world

Thank You