

Building universal understanding

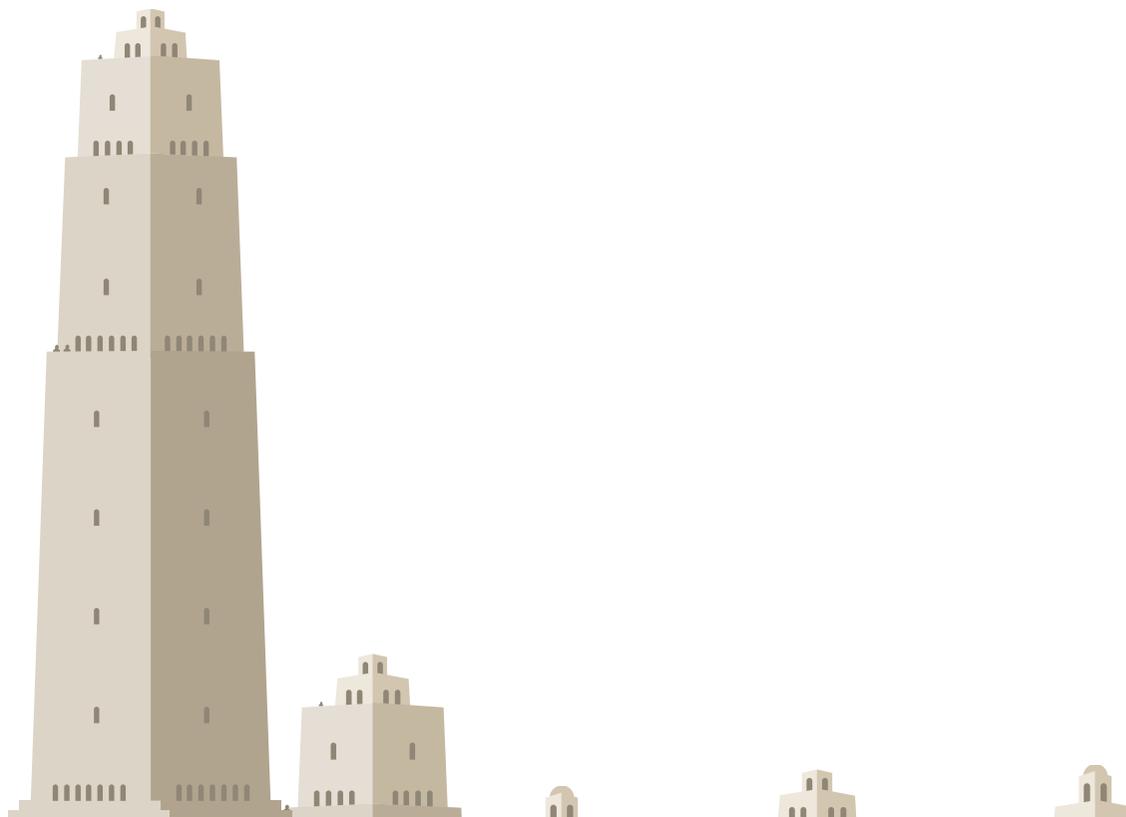
# Combining Crowd and AI to scale professional-quality translation

João Graça  
CTO



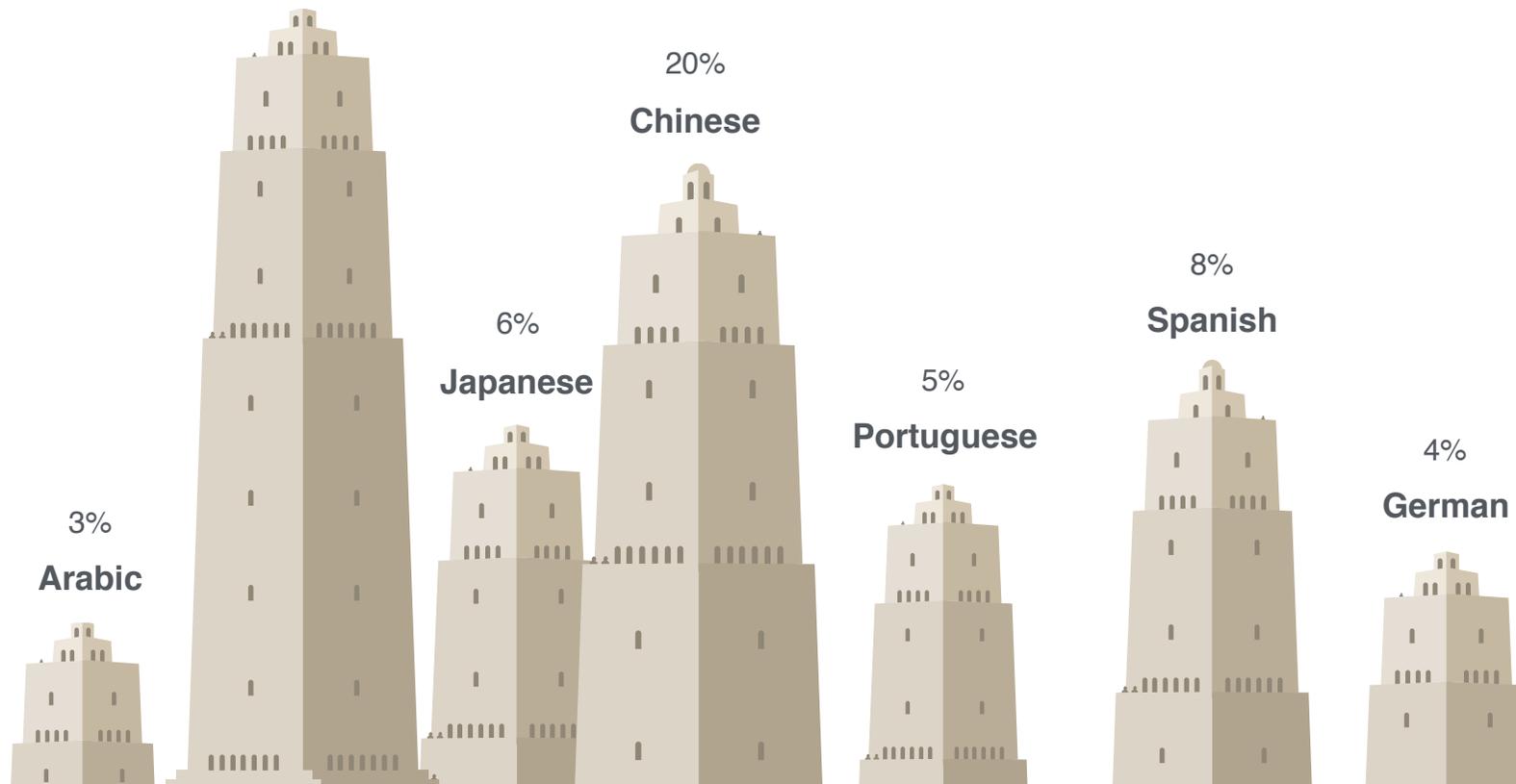
# The internet, 1997

80%  
English



# The internet, 2017

30%  
English



# Language barriers = trade barriers

*“Everyone  
speaks English”*  
costs the UK

**£48B**

3.5% UK GDP every year



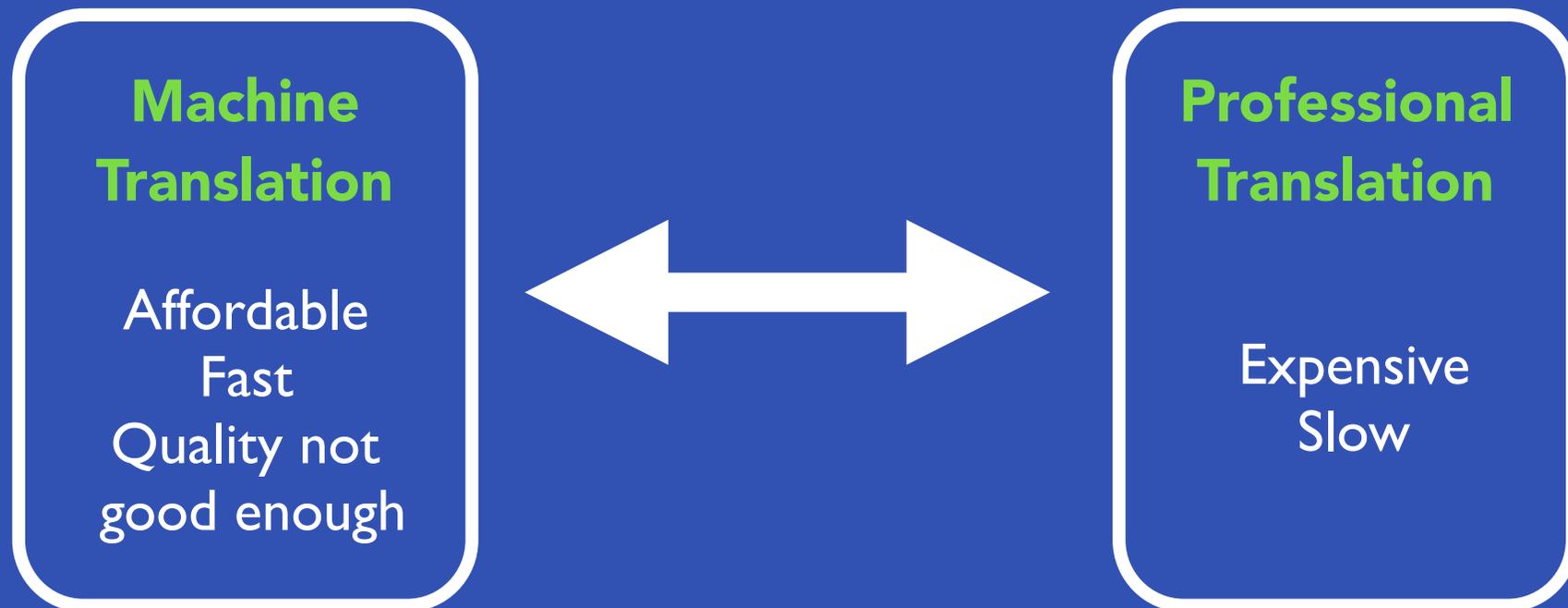
**Just 12%**  
of EU retailers **sell online**  
to other EU countries

**Just 15%**  
of EU consumers **buy online**  
from other EU countries



# Available Solutions

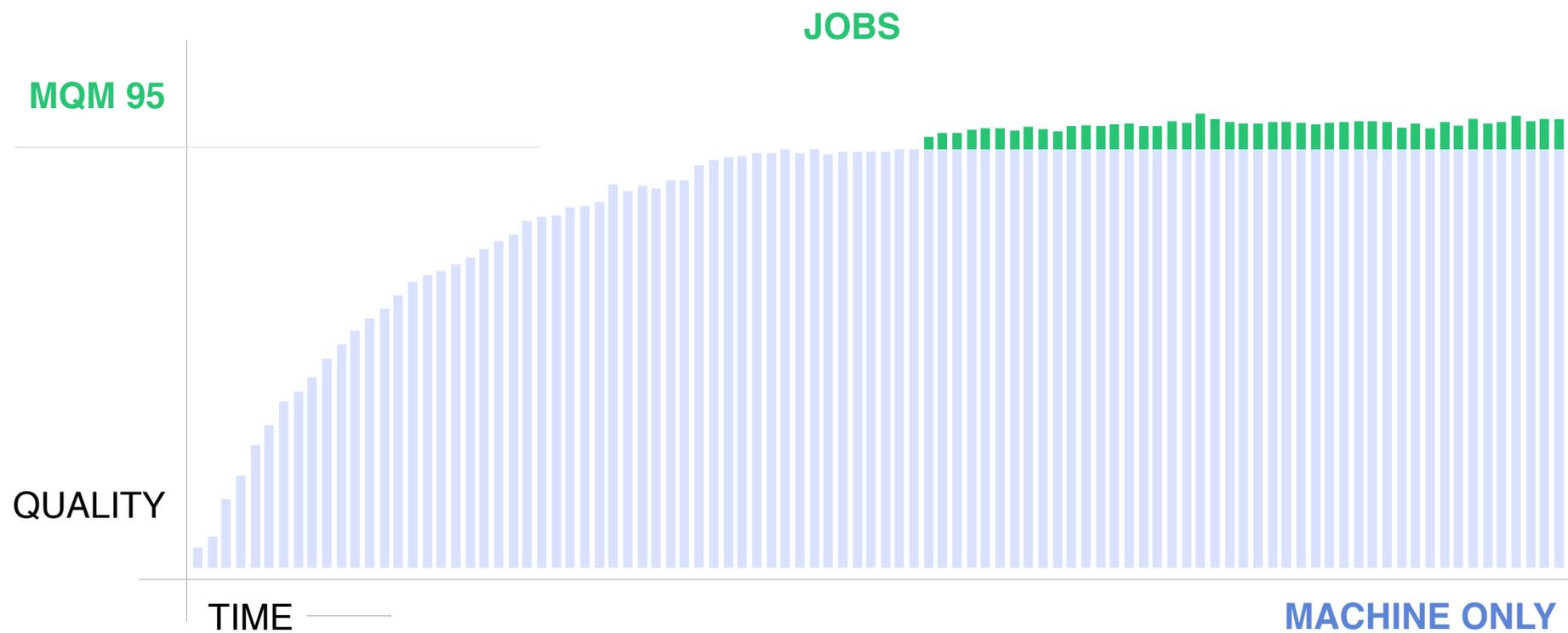
Lack of fast, affordable translation with human quality



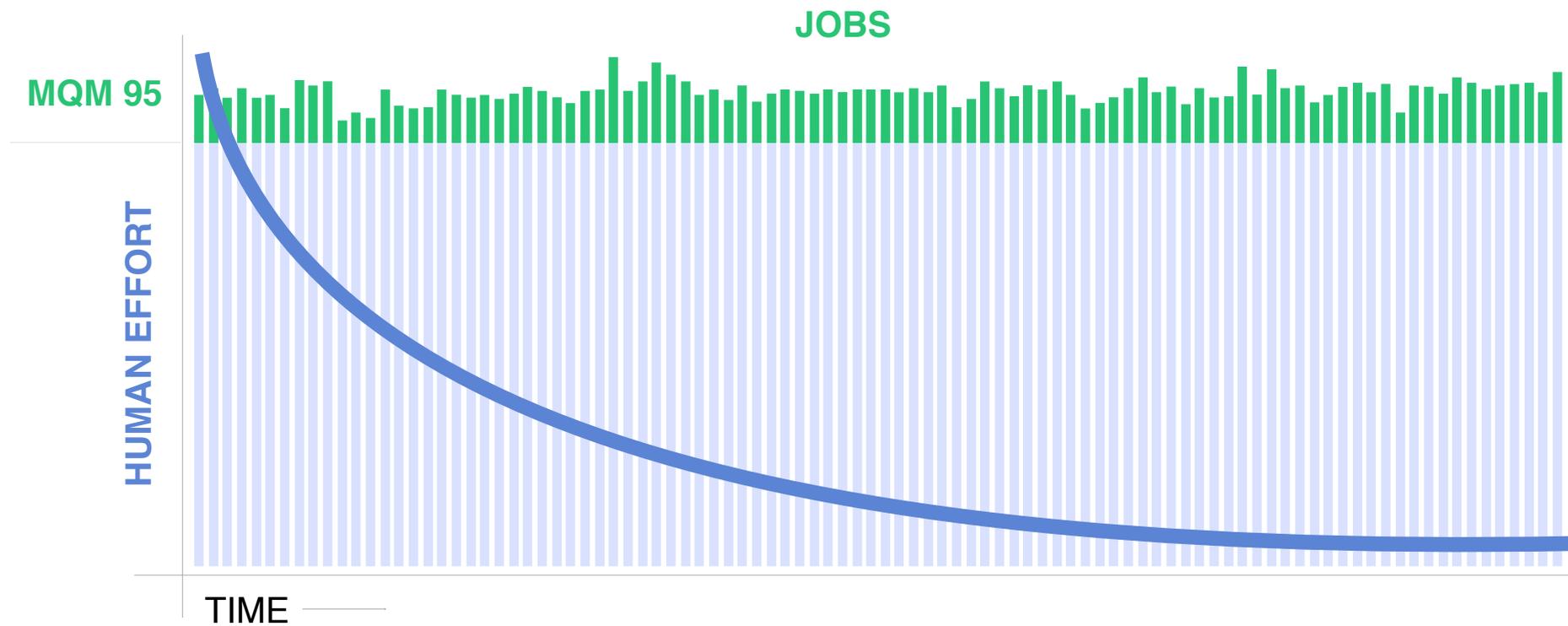
“All translation firms together are able to translate **far less than 1%** of relevant content produced everyday”

**CSA – MT Is Unavoidable to Keep Up with Content Volumes**

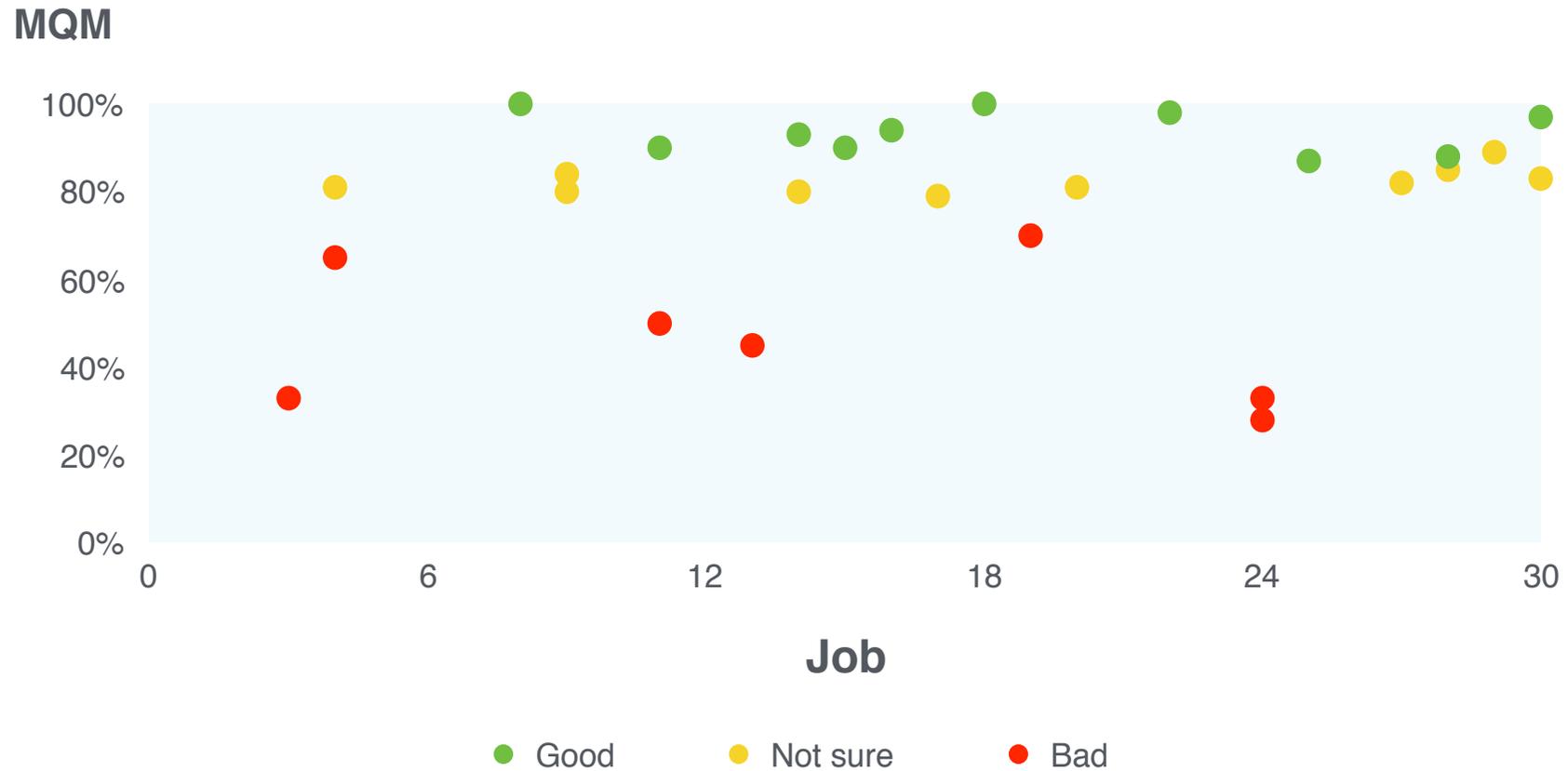
# Will AI solve translation?



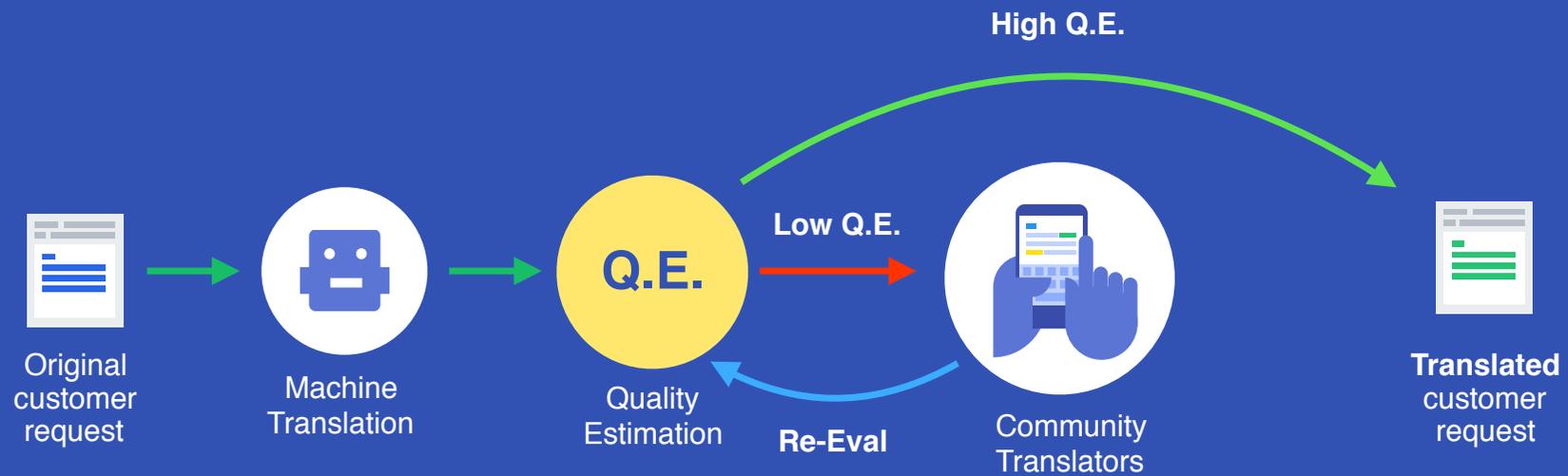
# Will AI solve translation?



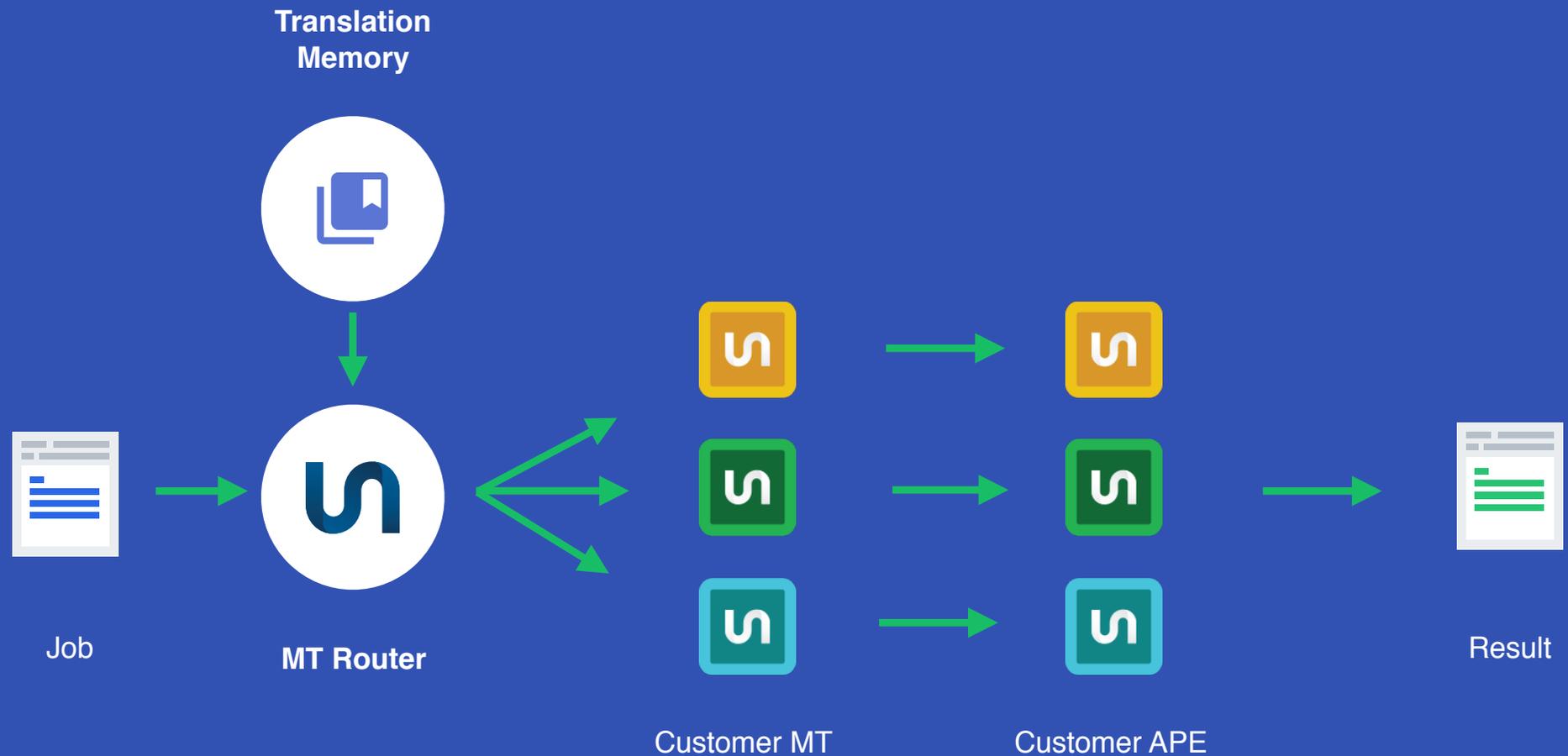
# Quality per Job



# Unbabel Pipeline



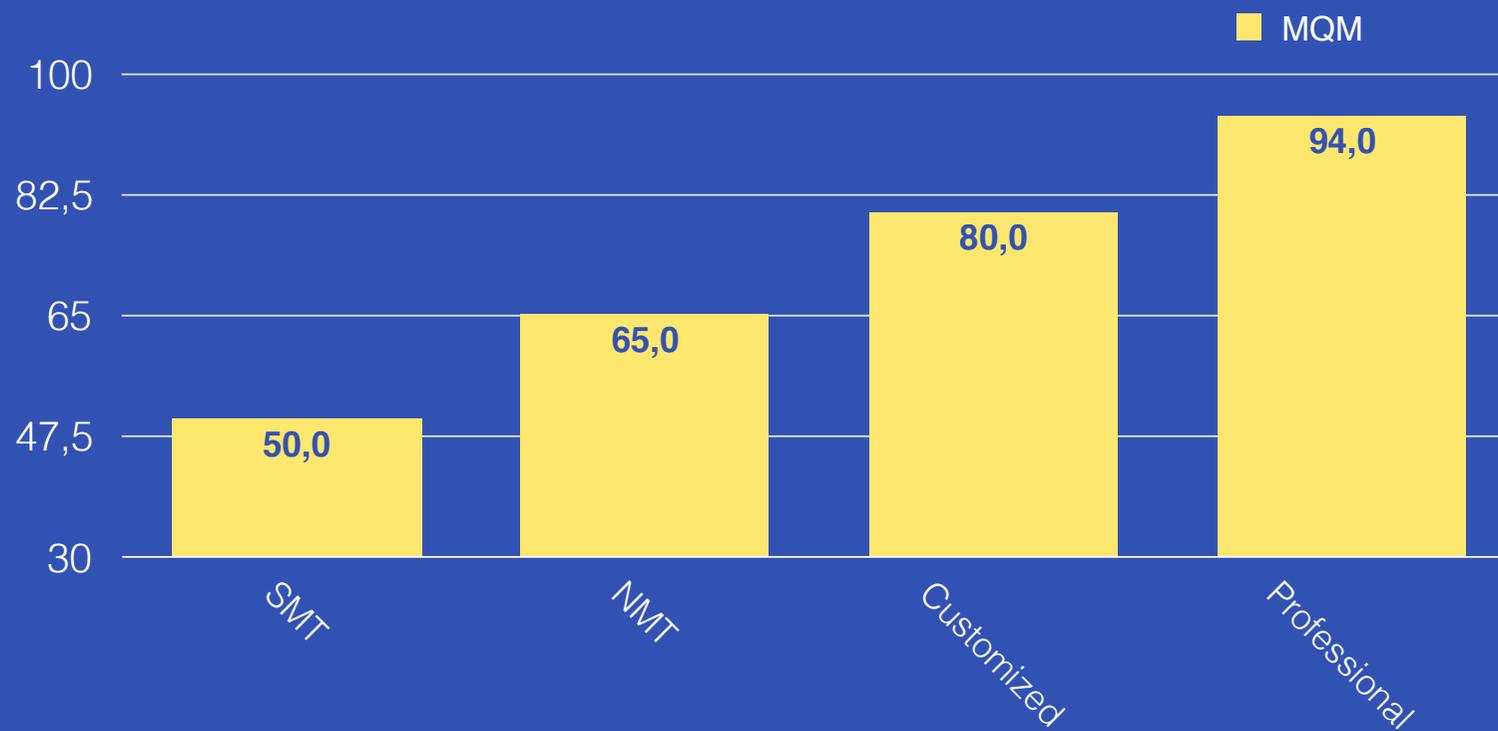
# Machine Translation Pipeline



# Customer Adaptation

## Customer Support Tickets

### MQM



# Quality Estimation



## Word-Level QE

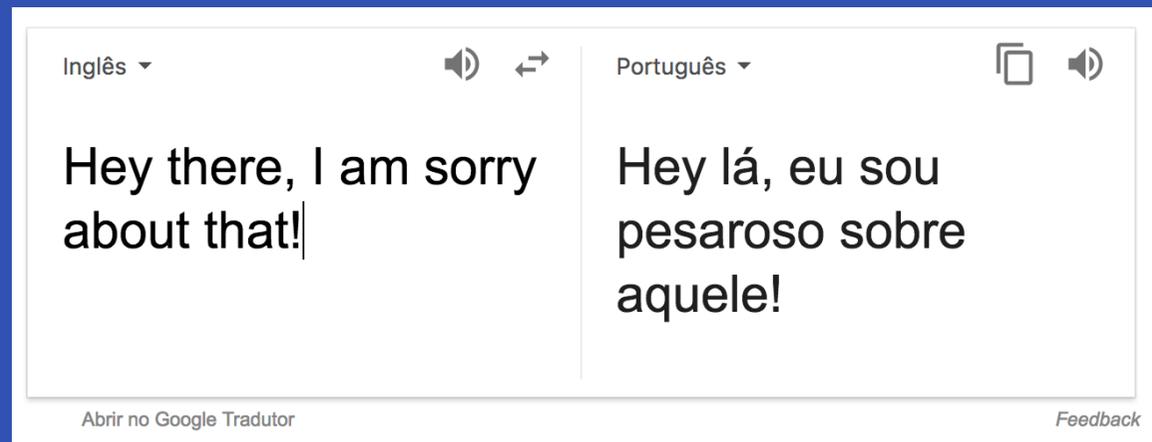
Which words are translated correctly/incorrectly?

## Sentence-Level QE

How good is the entire translation?

# Quality Estimation

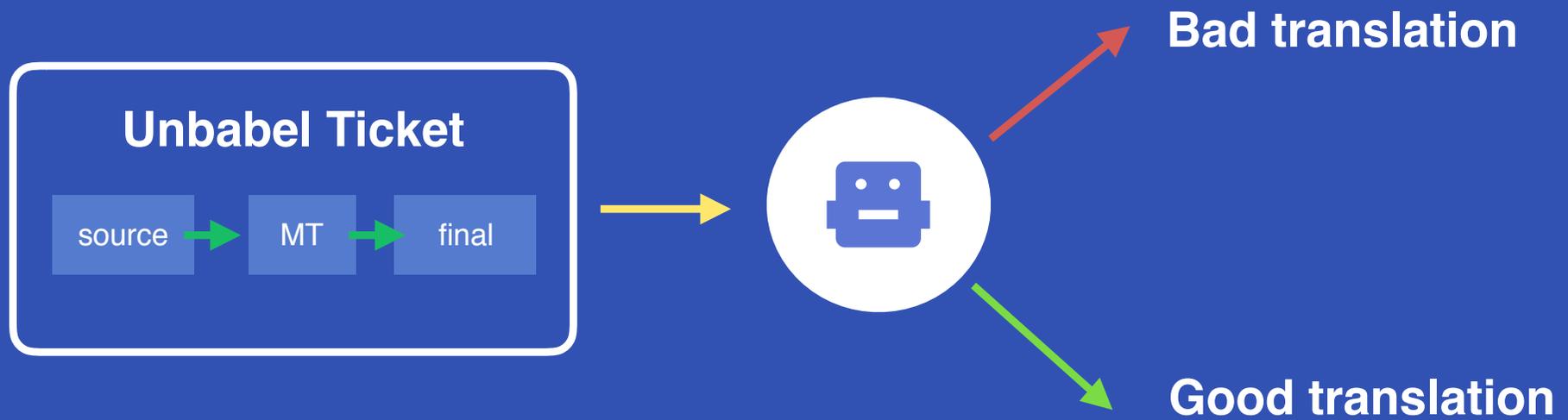
## Word-level QE example



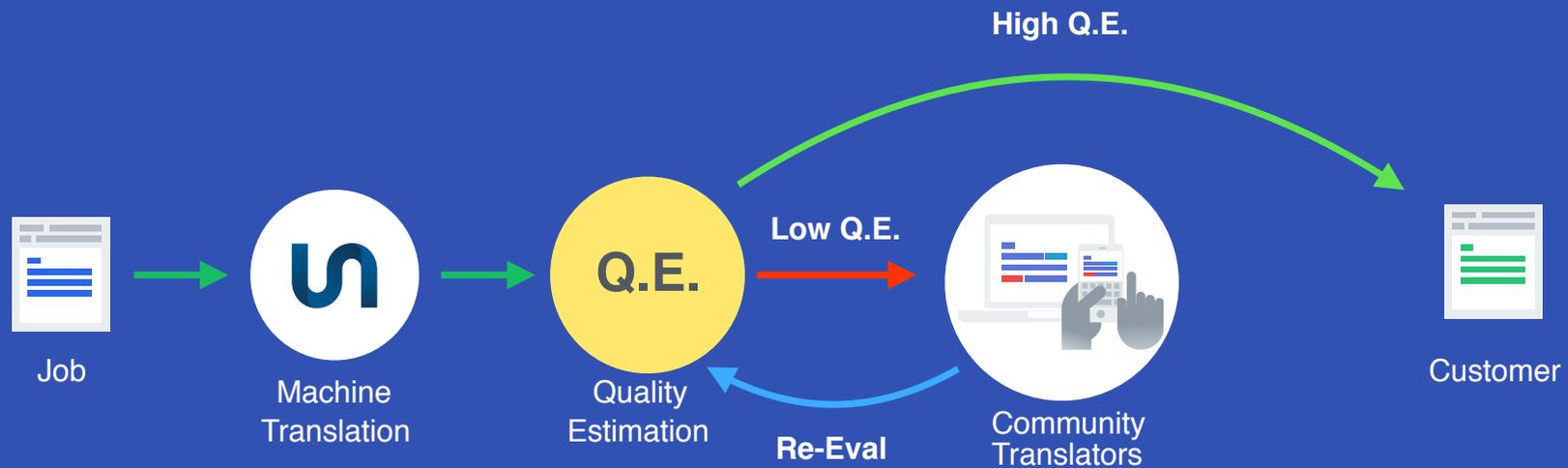
Hey lá , eu sou pesaroso sobre aquele !

BA BA OK BA BA BA  
D D D D D OK OK OK

# QE Training



# QE in the Pipeline



## Document-Level QE

how good is the entire document?

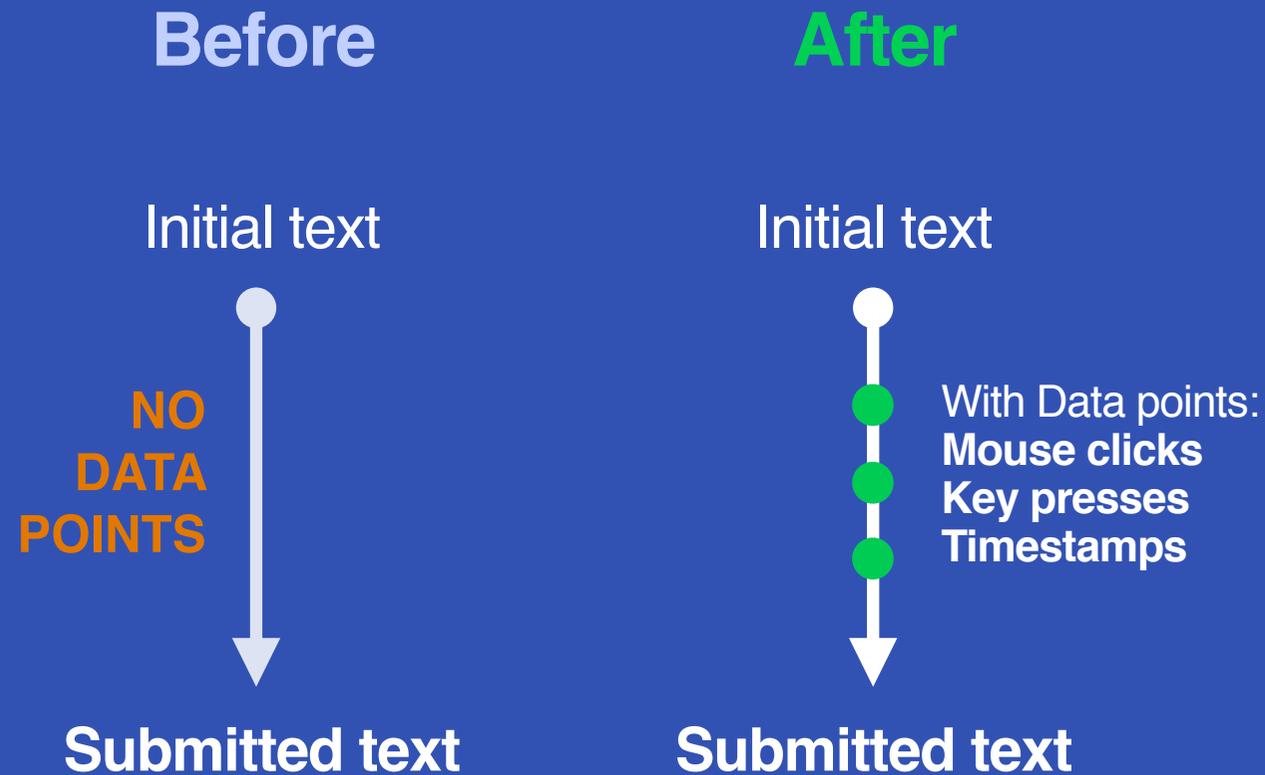
## Human QE

Can we evaluate post-edit output?

# Data Generation Engine



# Data Generation Engine



# Keystroke Analysis

## Raw data

At 18:03:30:  
In nugget 3  
mouseClick  
Cursor at 16  
Selected: 0  
At 18:03:31:  
In nugget 3  
Pressed Backspace  
Cursor at 16  
Selected: 0  
At 18:03:31:  
In nugget 3  
Pressed Backspace  
Cursor at 15  
Selected: 0  
At 18:03:31:  
In nugget 3  
Pressed Backspace  
Cursor at 14  
Selected: 0

At 18:03:35:  
In nugget 3  
Pressed Shift  
Cursor at 25  
Selected: 0  
At 18:03:35:  
In nugget 3  
Pressed s  
Cursor at 25  
Selected: 0  
At 18:03:35:  
In nugget 3  
Pressed i  
Cursor at 26  
Selected: 0  
At 18:03:35:  
In nugget 3  
Pressed e  
Cursor at 27  
Selected: 0

At 18:03:30:  
In nugget 3  
mouseClick  
Cursor at 16  
Selected: 0  
At 18:03:31:  
In nugget 3  
Pressed Backspace  
Cursor at 16  
Selected: 0  
At 18:03:31:  
In nugget 3  
Pressed Backspace  
Cursor at 15  
Selected: 0  
At 18:03:31:  
In nugget 3  
Pressed Backspace  
Cursor at 14  
Selected: 0

## Processed information

Initial text

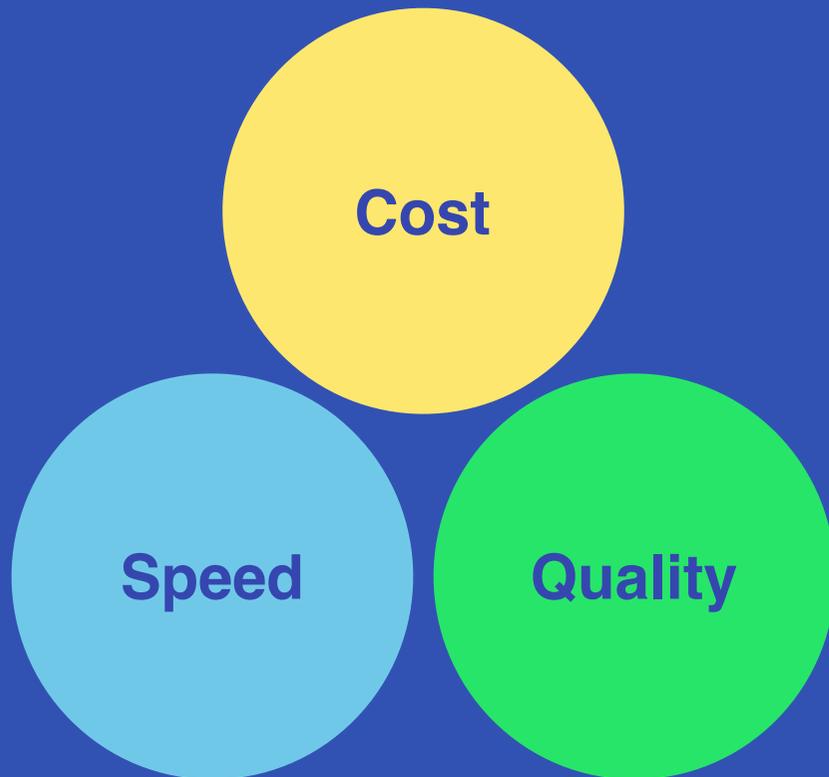
“Espero que esto es útil”

- Deleted word “es”
- Inserted word “sea”

Submitted text

“Espero que esto sea útil”

# Profession translation



## Unbabel pillars

- Editors Pool
- Initial Text (MT)
- Editor Assignment
- Custom Editing Interfaces
- Constant Quality Evaluation

# Unbabel Community



# Editors Pool

- 4 More specialization layers will be created
- 3 **Only the best rated editors have access to customer tasks**
- 2 Editors get rated with training tasks
- 1 First tests right after signup



# Evaluation Tool

The screenshot displays the 'Evaluation Tool' interface. At the top, it shows the language pair 'English to Spanish', the tone 'Formal', and the editor 'joel.sosa76'. Below this, a yellow box indicates 'No instructions available.' The main content area is divided into 'Source' and 'Translation' sections. The source text is 'Hi, Here is a support page with common credit card' and the translation is 'Hola: Aquí hay una página de soporte con problemas'. Below the translation, an 'Evaluation Report' is shown with a 'Final score: 3'. The report includes 'English to Spanish Linguist notes' and lists three observations: 'There were no edits done to the text. This results in a forced sentence structure.', '- "Hola" should be followed by a full stop.', and '- "Here is" does not necessarily mean "aquí hay", but rather "aquí tiene" or "le facilitamos".'

**Document Level Human QE**

# Deep Annotations

Job 1/25 (Annotator: ingrid.testa1, batch completed. Batch: HQ\_Batch\_20-08-2017\_to\_27-08-2017 QTZ1 Score: 105.93 Tone: Formal)

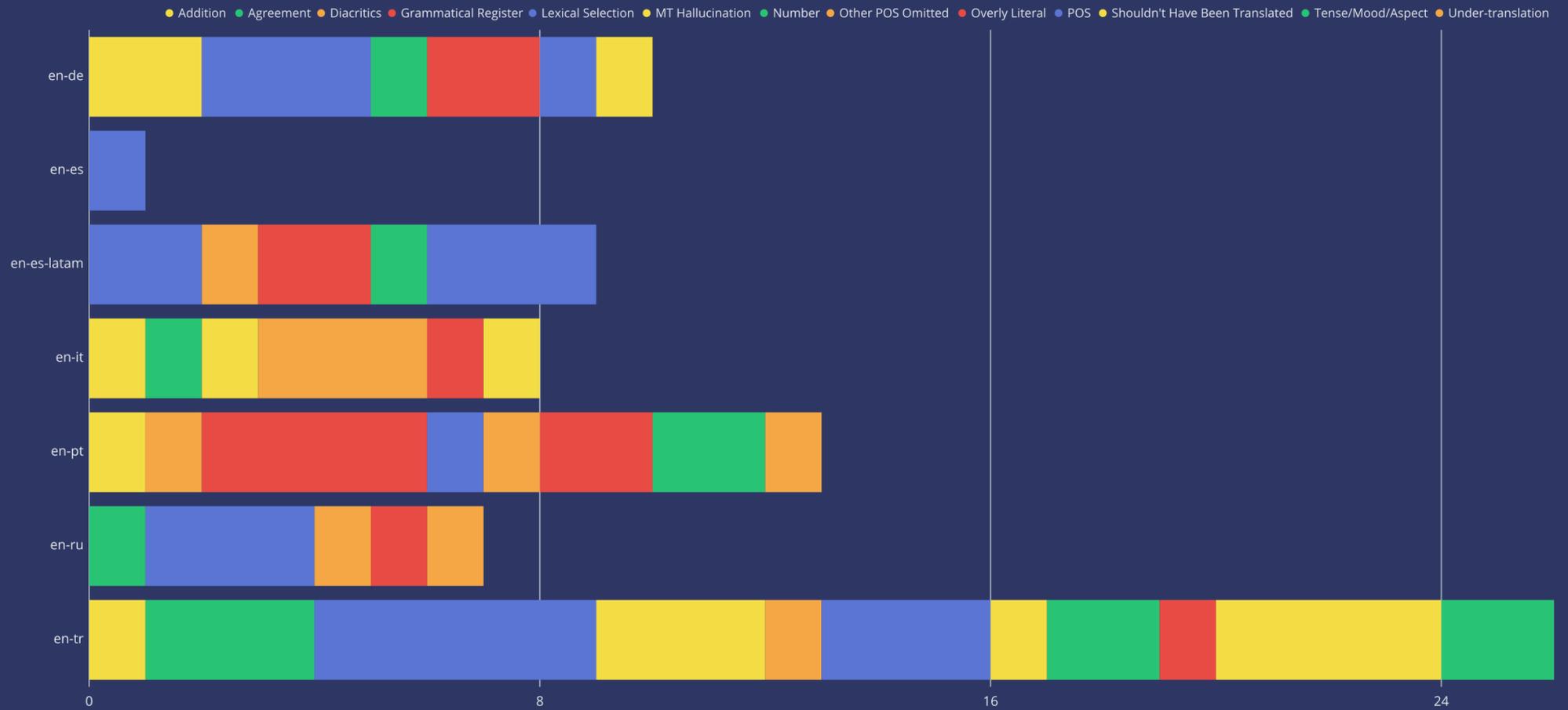
Client Instructions: These are Customer Support tickets for Rovio, game company that developed Angry Birds and others. Leave game titles and characters in English.

Original	Translation
Hello there,	Ciao,
and thanks for contacting us!	e grazie per averci contattato!
We're sorry to hear that the game is causing trouble.	Siamo spiacenti di sapere che il gioco le sta causando problemi.
In order to help you we need some additional information about the problem please.	Per aiutarti abbiamo bisogno di ulteriori informazioni sul problema per favore.
Which game are you referring to?	Quale gioco stai facendo riferimento?
What exact model is the device you are using and what operating system version does it have?	Quale modello esatto è il dispositivo che stai utilizzando e quale versione del sistema operativo ha?
What happens exactly when you are trying to play the game?	Cosa succede esattamente quando si sta cercando di giocare il gioco?
Do you get an error message?	Riceve un messaggio di errore?
Have you connected the game to your Rovio account or Facebook?	Hai connesso il gioco al tuo account Rovio o Facebook?
If yes, what's your login email address?	Se sì, qual è il tuo indirizzo di posta elettronica di accesso?
We apologize for any inconvenience and look forward to your response	Ci scusiamo per gli eventuali disagi e restiamo in attesa di una sua risposta

Annotations	
Type	Context
Register	Scopri
Register	aiutarti
Register	Hai
Overly Literal	Cosa succede esattamente quando si sta cercando di giocare il gioco
Word Order	Rovio Newsletter
Prepositions	
Overly Literal	Quale modello esatto è il dispositivo
Register	Ciao
Untranslated	Support Team
Register	tuo
Register	stai
Register	stai
Register	tuo

# Error Analysis

Critical Errors MT



# QE for Annotation

The screenshot shows the Annotate web interface. At the top, the logo "Annotate. powered by Unbabel." is on the left, and "Admin", "Statistics", and "Hi Annotatoradmin" are on the right. Below the header, a job overview bar displays "Job 1/25 (Annotator: Ingrid.testa1, batch completed. Batch: HQ\_Batch\_20-08-2017\_to\_27-08-2017 QTZ1 Score: 105.93 Tone: Formal)".

Client Instructions: These are Customer Support tickets for Rovio, game company that developed Angry Birds and others. Leave game titles and characters in English.

Annotations

Type	Context
Register	stai
Register	stai
Register	tuo

operating system version does it have? quale versione del sistema operativo ha?

What happens exactly when you are trying to play the game? Cosa succede esattamente quando si sta cercando di giocare il gioco?

Do you get an error message? Riceve un messaggio di errore?

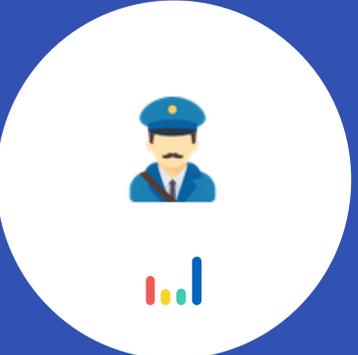
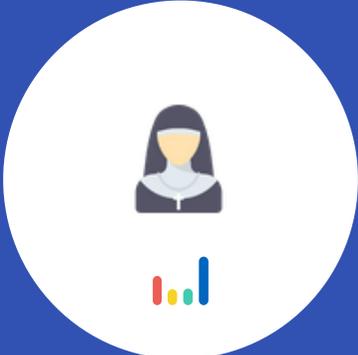
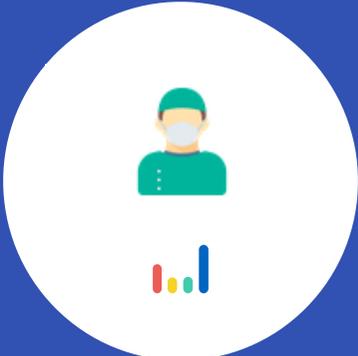
Have you connected the game to your Rovio account or Facebook? Hai connesso il gioco al tuo account Rovio o Facebook?

If yes, what's your login email address? Se sì, qual è il tuo indirizzo di posta elettronica di accesso?

We apologize for any inconvenience and look forward to your response. Ci scusiamo per gli eventuali disagi e restiamo in attesa di una sua risposta.

## Pre-fill with word level QE

# Editors Profiling



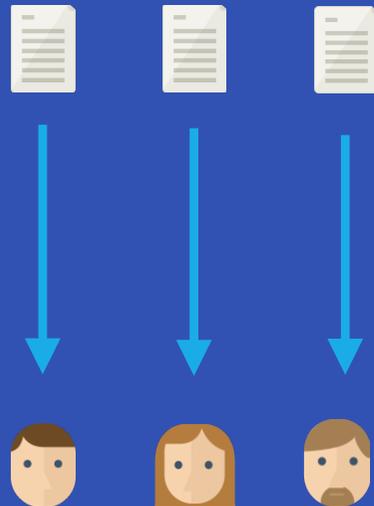
# Editor Assignment

Queue	Topics	Priority	SLA	Tasks/time	Editors	Rating	Native	Topics
G		1000	6 H	 2 m		4.2		
G		1100	30 m	 6 m		3.8		
G		1000	2 D	 10 m				
G		1000	6 D	 12 m		4.3		
R		1100	20 m	 18 m		4.8		
R		1100	40 m	 45 m				

← Pull

# Editor Assignment

Regular distribution



★★★★★ 3.8

old rating

Smart distribution



★★★★★ 4.6

Improved rating

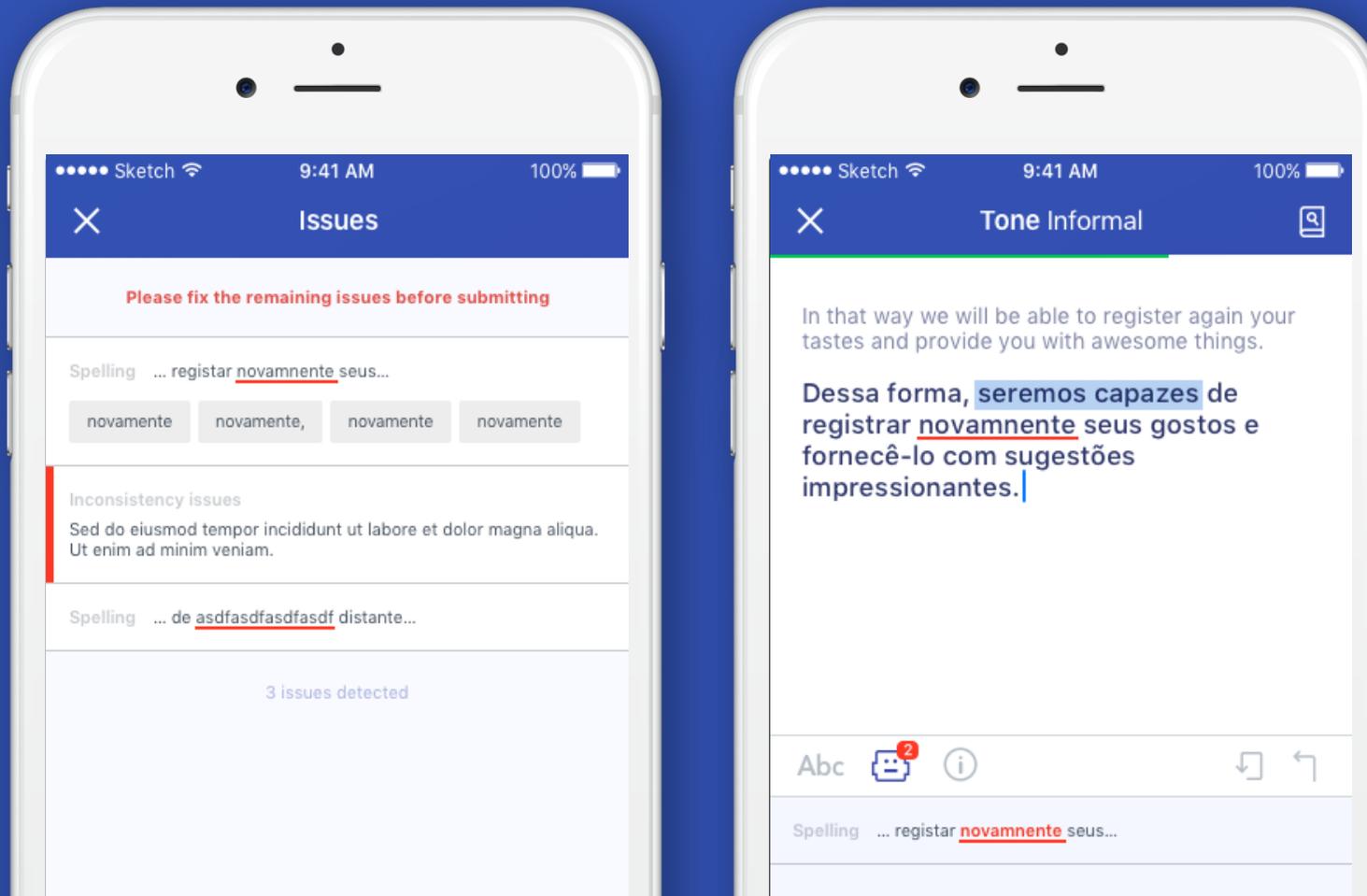
# Post-Editing Interfaces

The screenshot shows a web browser window with the URL `https://web-interface-demo.unbabel.com/editor/paid_tasks/?lp=en_pt&filename=5`. The page title is "Paid Task | English to Spanish • 46 words". The interface is split into two main columns: "Original" (English) and "Translation" (Spanish). The original text reads: "This article assumes that you understand how to create a WordPress plug-in. If you don't, read the WordPress Codex article Writing a Plugin before continuing. For my example, I used the WP-CLI to create a scaffold for my plug-in from the command line." The translation reads: "Este artículo asume que usted comprende cómo crear un complemento de WordPress. Si no lo hace, lea el artículo de WordPress Codex Escribir un Plugin Antes de continuar. Para mi ejemplo, utilicé el WP-CLI para crear un complemento desde la línea de". A blue box highlights the word "ejemplo" in the translation, and a context menu is open over it, showing suggestions: "ejemplo", "empalo", "empelo", "empilo", "Save to dictionary", and "Ignore". The menu also indicates "Misspelled word: ejemplo". On the right side, there are tabs for "INSTRUCTIONS", "DICTION...", and "SMARTC...". Below these, there is a "Register" section with "Formal" and a "Customer Brief" section with "Long Spanish Formal". At the bottom of the main editor area, there are "SKIP AND LEAVE" and "SUBMIT" buttons.

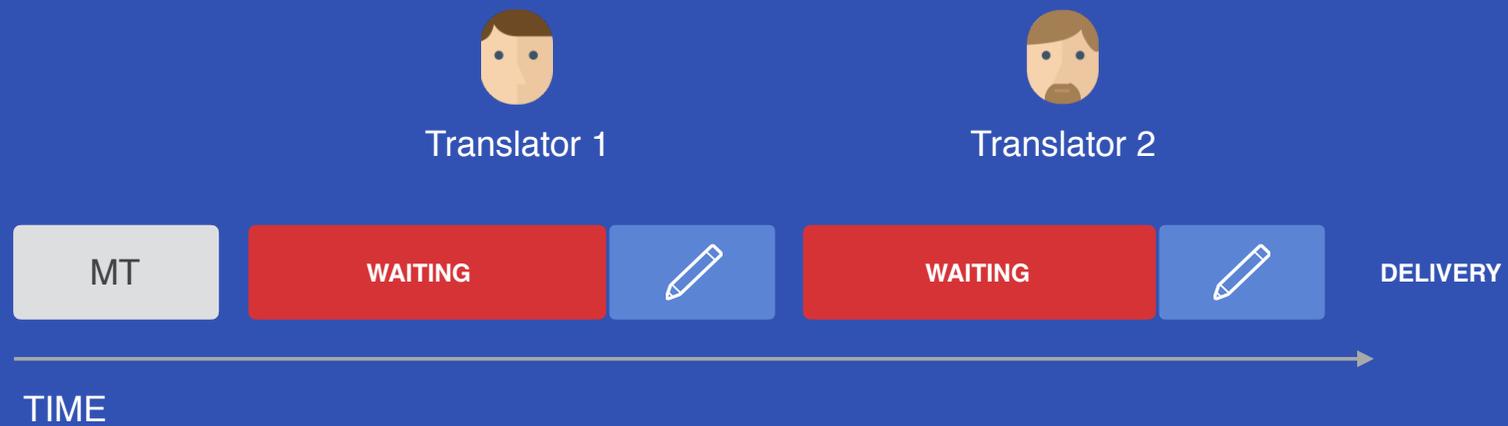
# QE on Interfaces

The screenshot displays a web browser window with the URL `https://web-interface-demo.unbabel.com/editor/paid_tasks/?lp=en_pt&filename=4`. The page title is "Paid Task | English to Portuguese • 162 words". The interface is split into two main columns: "Original" (English) and "Translation" (Portuguese). The original text is: "Hi Cesar, I'm sorry to hear about this misunderstanding! You can click 'Menu' and change your subscription or visit the 'Subscriptions' page here: URL-0 The unused value of your current plan will be applied to the cost of your new plan. Regarding the price, Cambly Private gives you one-on-one tutoring with a native English speaker. For this type of service, we are Please just let us know how you'd like to proceed. Best, Noreen". The translated text is: "Olá Cesar, Estou triste de ouvir sobre este mal-entendido! Você pode clicar em 'Menu' e mudar sua assinatura ou visite a página 'Assinaturas' aqui: URL-0 O valor não utilizado de seu plano atual será aplicada ao custo de seu novo plano. Quanto ao preço, Cambly Privada lhe dá um-em-um tutoria com um falante nativo Inglês. Para este tipo de serviço, estamos realmente com preços extremamente competitivos. Por favor, deixe-nos saber como você gostaria de prosseguir. melhor, Noreen". A green checkmark and a tooltip are visible over the translation. The tooltip text is: "High Quality Segment Approved by the quality system. Please edit only if necessary." On the right side, there is a sidebar with a "Customer Brief" section containing instructions: "This task contains text that was hidden for privacy. Do not delete. This is a customer support message from Skyscanner. You might sometimes be dealing with complaints so please always use a FRIENDLY FORMAL TONE (using formal pronouns to address the reader but a style and vocabulary close to informal). For SPANISH and SPANISH LATAM, please use INFORMAL PRONOUNS. Do not translate names, airlines and websites."

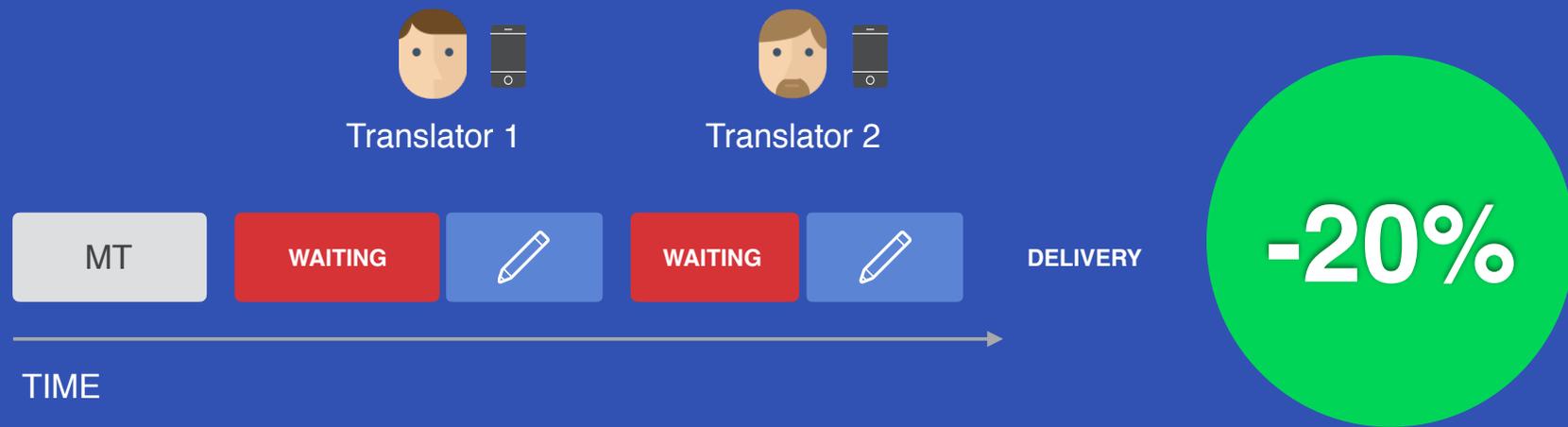
# Post-Editing Interfaces



# Time Spent on Job



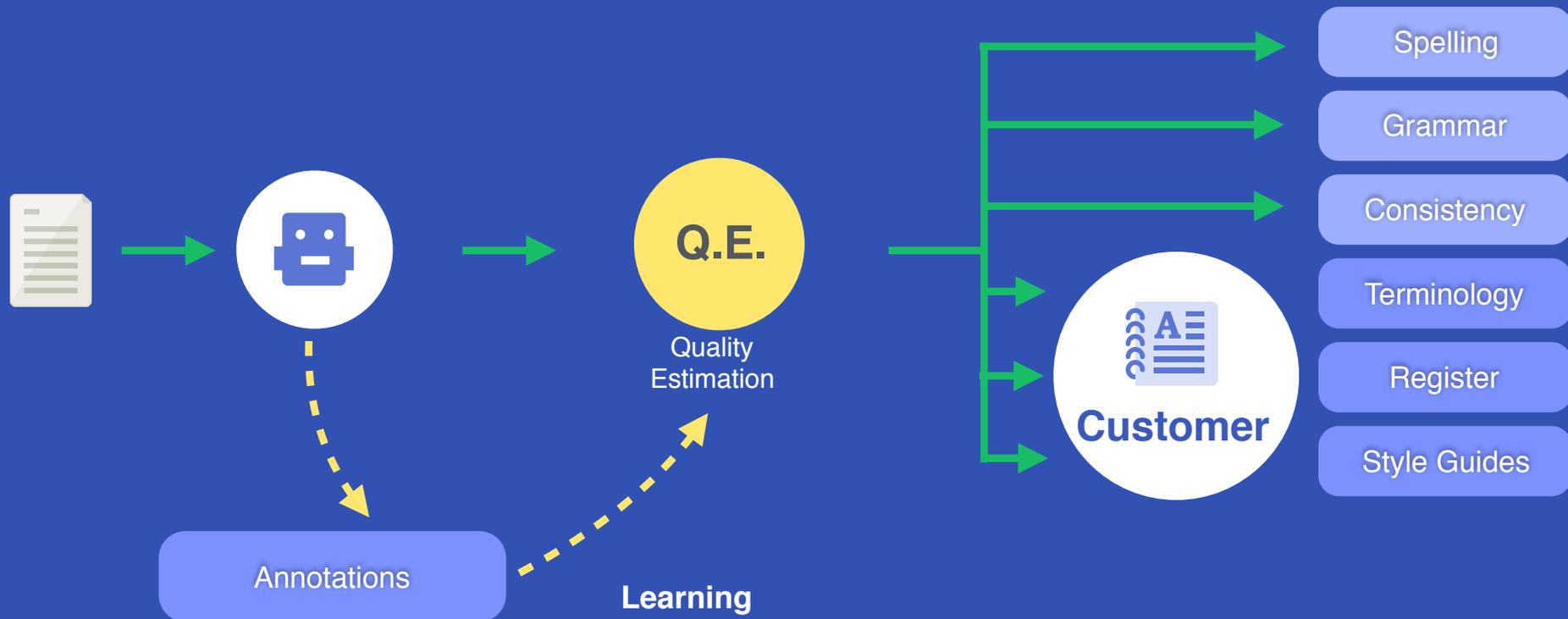
# Time Spent on Job: Mobile

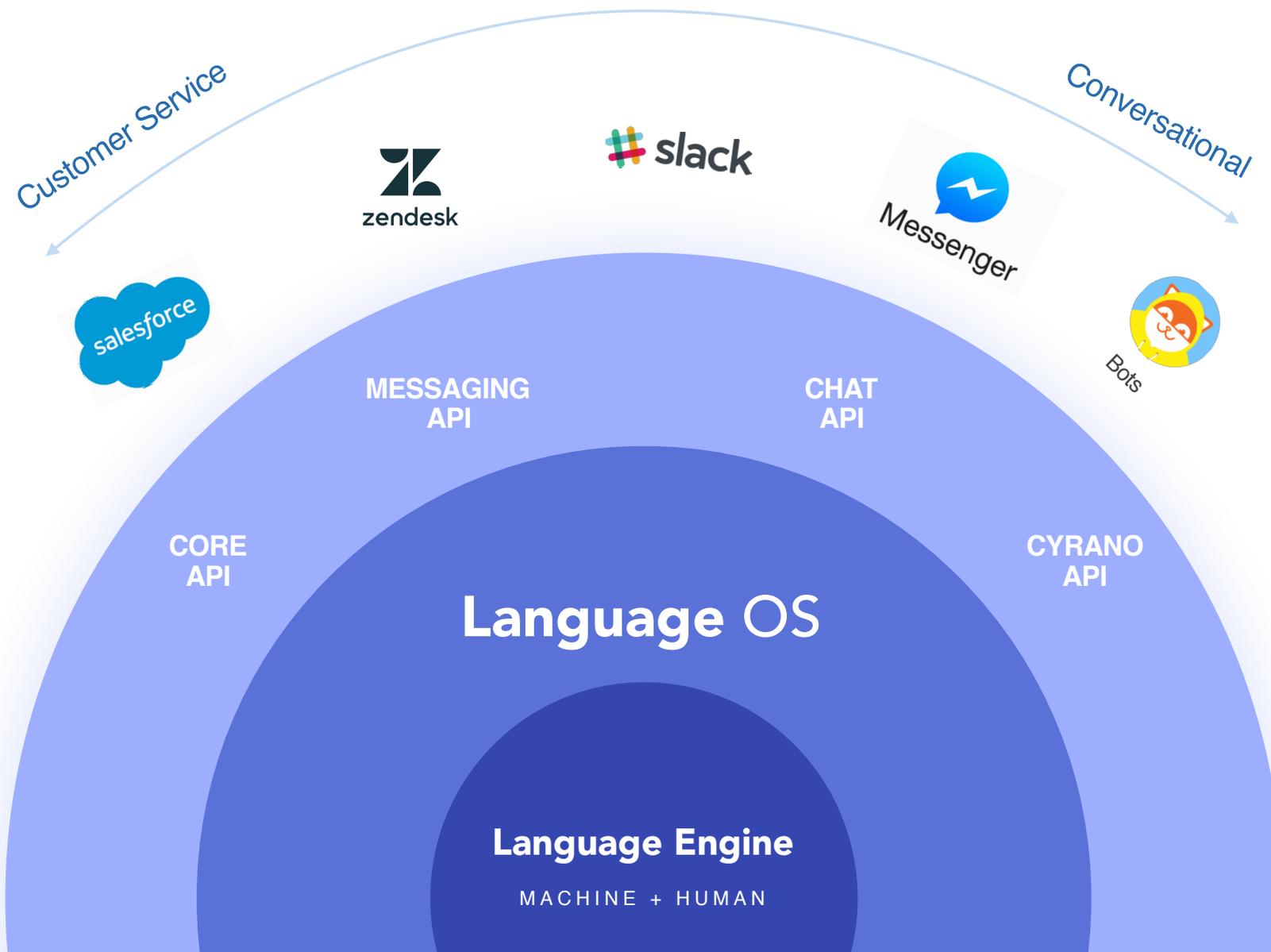


# Smartcheck

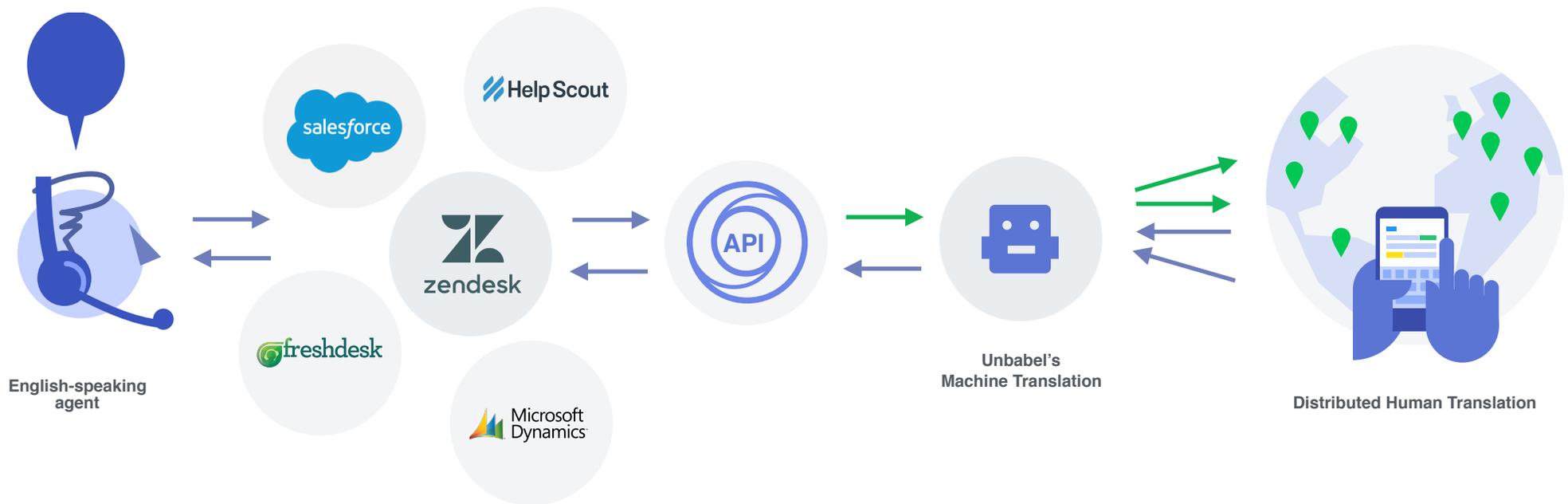


# Smartcheck (QE Version)





# Unbabel for Customer Service



**Unbabel adapts to any workflow**

# Customer Replies: **Speed & Quality**



20 minutes



94

# Unbabel Chat

Native speaking  
in multiple languages



# Chat



## Unbabel Chat

Understand and be understood in multiple languages.

with Juan Torres 

Start typing here, or paste the text you'd like to be translated.

Translate to

Portuguese

Translate

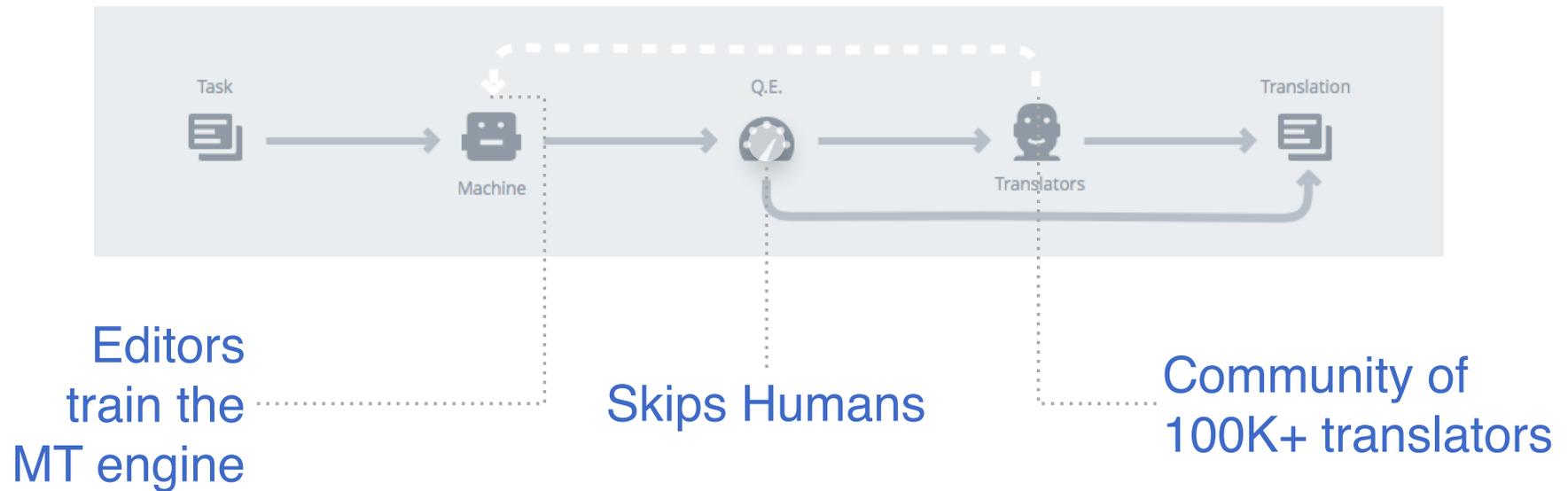
Good morning, that's great! What color do you like?

 Translating...

Good morning, I would like to buy a hat

Bom dia, gostava de comprar um chapéu

# Chat Translation Flow



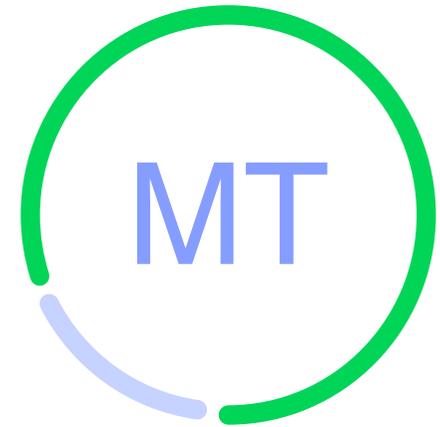
# Chat Messages: **Speed & Quality**



**2 minutes**

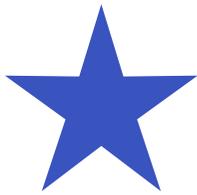


**90**

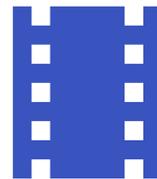


**80%**

# Other Use Cases



Reviews



Video



Newsletters



Travel  
descriptions



SEO



# We're Hiring

<https://unbabel.com/careers/>

