



# PARTNERSHIPS FOR MACHINE TRANSLATION

JOHN DI RICO - @JOHNNYDIRICO

APEXTRA.NET / WORDFAST.COM

# OUTLINE



Background



Hypotheses



Methods

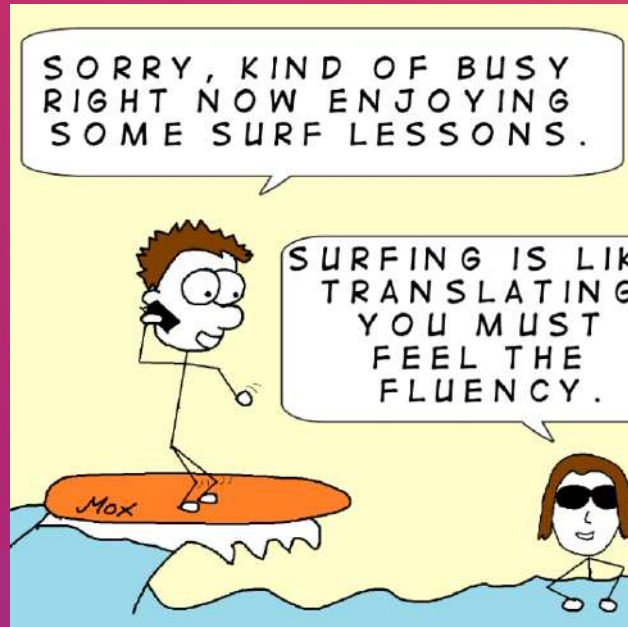


Results



Conclusion

# BACKGROUND



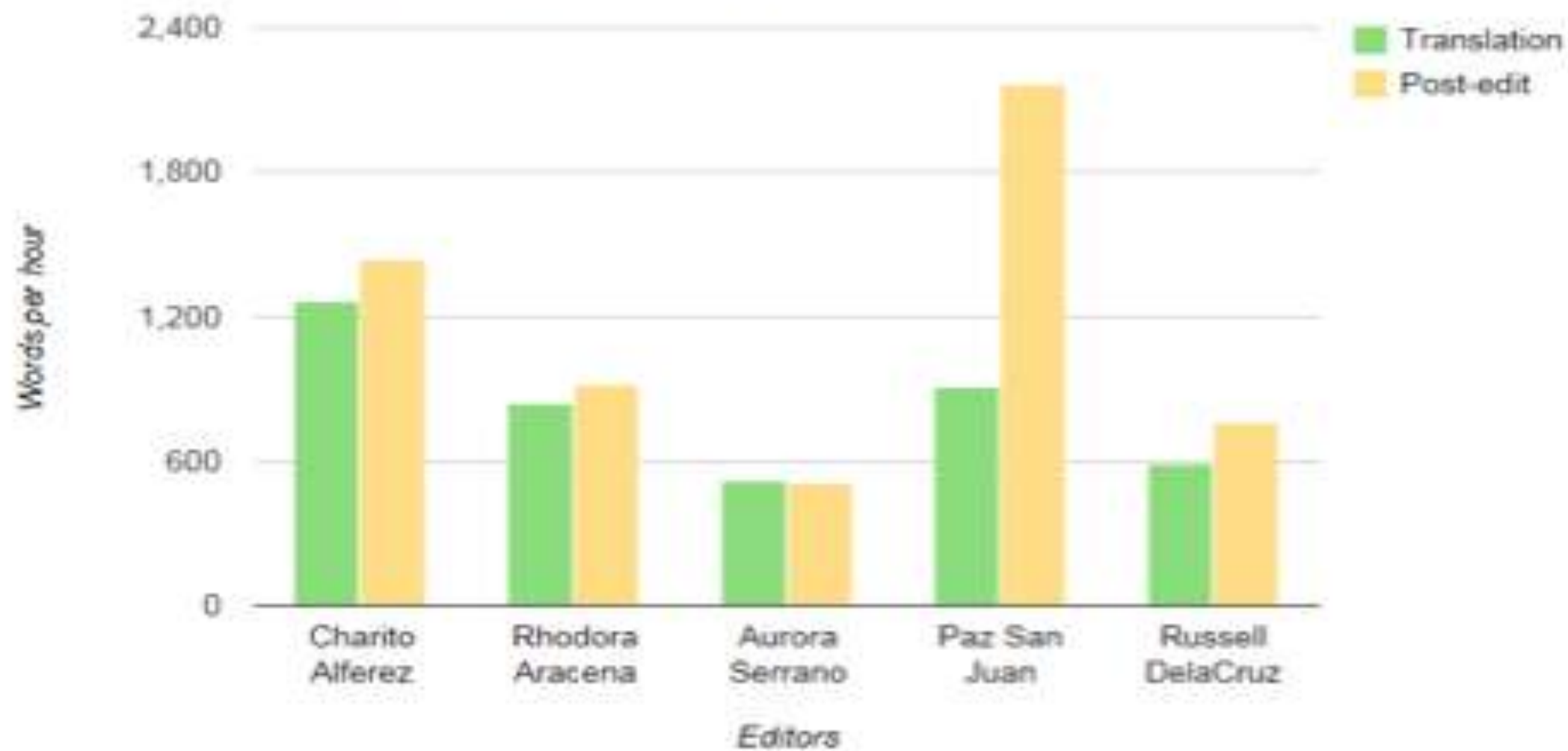
TRANSLATION COMPANY, TOOL PROVIDER AND TRAINER

## Productivity per Editor

Evaluator Name	Translation (WPH)	Post-Edit (WPH)	Time Spent for Translation (seconds)	Time Spent for Post-edit (seconds)	Post-Editing Productivity Ratio <a href="#">(Info)</a>
Charito Alferez	1,265	1,442	3738	3409	<b>1.13</b>
Rhodora Aracena	840	915	5630	5372	<b>1.08</b>
Aurora Serrano	519	514	9100	9564	<b>0.99</b>
Paz San Juan	906	2,168	5216	2267	<b>2.39</b>
Russell DelaCruz	588	760	8034	6467	<b>1.29</b>

**2014... EFFICIENCY UP 50%**

### English (United States) > Tagalog



# EFFICIENCY

- Increase efficiency without sacrificing quality
- How?
  - Translation memory
  - QA Tools
  - Post-editing MT
  - Voice recognition
  - Hybrid voice / MT
  - ...





## **A Computer Journal For Translation Professionals**

**Issue 21-1-321**

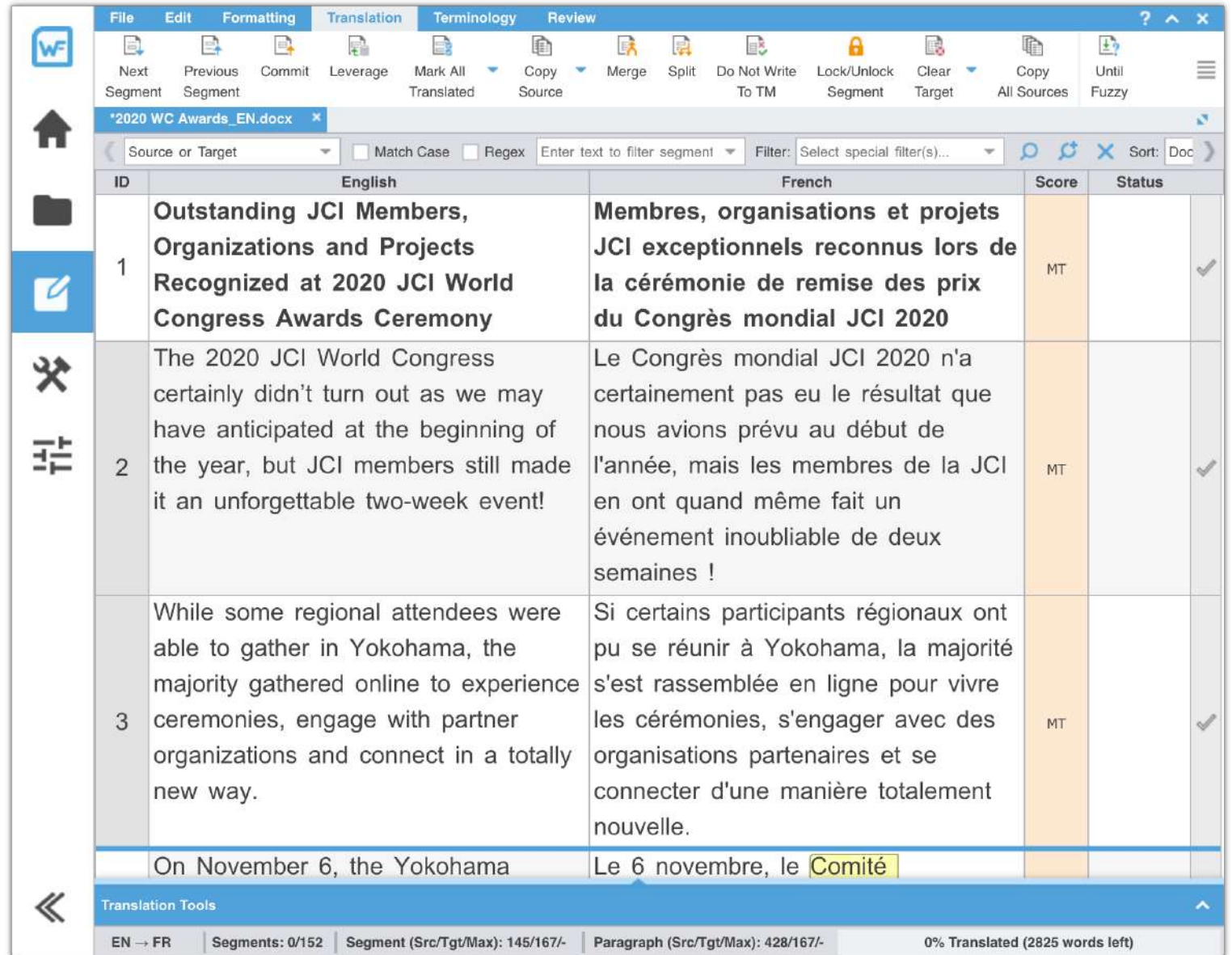
(the three hundred twenty first edition)

**JOST ZETZSCHE**

Finally, and I feel this might be the most important of my self-assigned tasks, I will continue to find and communicate good ways to work with machine translation. There has rarely been a *Tool Box Journal* issue where I have not at least mentioned that there are better ways of dealing with machine translation suggestions than traditional post-editing. I have tried to push the simultaneous use of several sources for machine translation suggestions as a pure repository of short and desired fragments, but there certainly are other ways -- and I know (I KNOW!) we're not working hard enough to find those. If the *Tool Box Journal* can be an instigator or a forum to find a wide range of optimized ways to use machine translation, or if it can serve as a lobby to change existing technology to work with those ways -- man! I can't tell you how excited I would be about that!



# POST-EDITING



The screenshot displays a translation software interface with the following components:

- Menu Bar:** File, Edit, Formatting, Translation, Terminology, Review.
- Toolbar:** Next Segment, Previous Segment, Commit, Leverage, Mark All Translated, Copy Source, Merge, Split, Do Not Write To TM, Lock/Unlock Segment, Clear Target, Copy All Sources, Until Fuzzy.
- Search Bar:** Source or Target, Match Case, Regex, Enter text to filter segment, Filter: Select special filter(s)...
- Table:**

ID	English	French	Score	Status
1	<b>Outstanding JCI Members, Organizations and Projects Recognized at 2020 JCI World Congress Awards Ceremony</b>	<b>Membres, organisations et projets JCI exceptionnels reconnus lors de la cérémonie de remise des prix du Congrès mondial JCI 2020</b>	MT	✓
2	The 2020 JCI World Congress certainly didn't turn out as we may have anticipated at the beginning of the year, but JCI members still made it an unforgettable two-week event!	Le Congrès mondial JCI 2020 n'a certainement pas eu le résultat que nous avons prévu au début de l'année, mais les membres de la JCI en ont quand même fait un événement inoubliable de deux semaines !	MT	✓
3	While some regional attendees were able to gather in Yokohama, the majority gathered online to experience ceremonies, engage with partner organizations and connect in a totally new way.	Si certains participants régionaux ont pu se réunir à Yokohama, la majorité s'est rassemblée en ligne pour vivre les cérémonies, s'engager avec des organisations partenaires et se connecter d'une manière totalement nouvelle.	MT	✓
- Status Bar:** EN → FR | Segments: 0/152 | Segment (Src/Tgt/Max): 145/167/- | Paragraph (Src/Tgt/Max): 428/167/- | 0% Translated (2825 words left)

# DISCERNMENT

The screenshot displays a translation software interface with the following components:

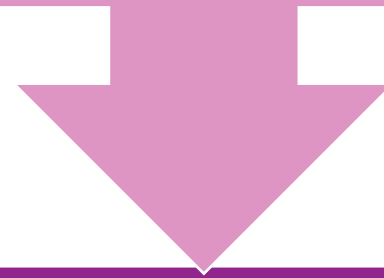
- Menu Bar:** File, Edit, Formatting, Translation, Terminology, Review.
- Toolbar:** Next Segment, Previous Segment, Commit, Leverage, Mark All Translated, Copy Source, Merge, Split, Do Not Write To TM, Lock/Unlock Segment, Clear Target, Copy All Sources, Until Fuzzy.
- Search Bar:** Source or Target, Match Case, Regex, Enter text to filter segment, Filter: Select special filter(s)...
- Main Table:**

ID	English	French	Score	Status
1	Outstanding <b>JCI Members</b> , Organizations and <b>Projects</b> Recognized at 2020 <b>JCI World</b> <b>Congress Awards Ceremony</b>	Membres, organisations et projets exceptionnels de la JCI reconnus lors de la cérémonie de remise des prix du <b>Congrès mondial JCI</b> 2020	MT	✓
2	The 2020 JCI World Congress certainly didn't turn out as we may have anticipated at the beginning of the year, but JCI members still made it an unforgettable two-week event! While some regional attendees were		-	✓
- TM Lookup Table:**

#	Source	Target	Score
1	Outstanding JCI Members, Organizations and Projects Recognized at 2020 JCI World Congress Awards Ceremony	Membres, organisations et projets exceptionnels de la JCI reconnus lors de la cérémonie de remise des prix du Congrès mondial JCI 2020	MT
2	Outstanding JCI Members, Organizations and Projects Recognized at 2020 JCI World Congress Awards Ceremony	Membres, Organisations et Projets exceptionnels reconnus lors de la Cérémonie de Récompenses du Congrès Mondial JCI 2020	MT
3	Outstanding JCI Members, Organizations and Projects Recognized at 2020 JCI World Congress Awards Ceremony	Membres, organisations et projets JCI exceptionnels reconnus lors de la cérémonie de remise des prix du Congrès mondial JCI 2020	MT
- Status Bar:** EN → FR | Segments: 0/152 | Segment (Src/Tgt/Max): 105/134/- | Paragraph (Src/Tgt/Max): 105/134/- | 0% Translated (2825 words left)

## HYPOTHESES

Most freelance translators have adopted some form of machine translation in their workflow.



Most of these translators use a single-source of machine translation.

# METHODS



Survey mail merged to 355 translators



Language pairs:

DE>EN (10), EN>ZH (3), EN>DE (60), EN>ES (9),  
EN>FR (65), EN>PO (4), EN>JA (33), FR>DE (61),  
IT>FR (46), FR>EN (53), Other (11)



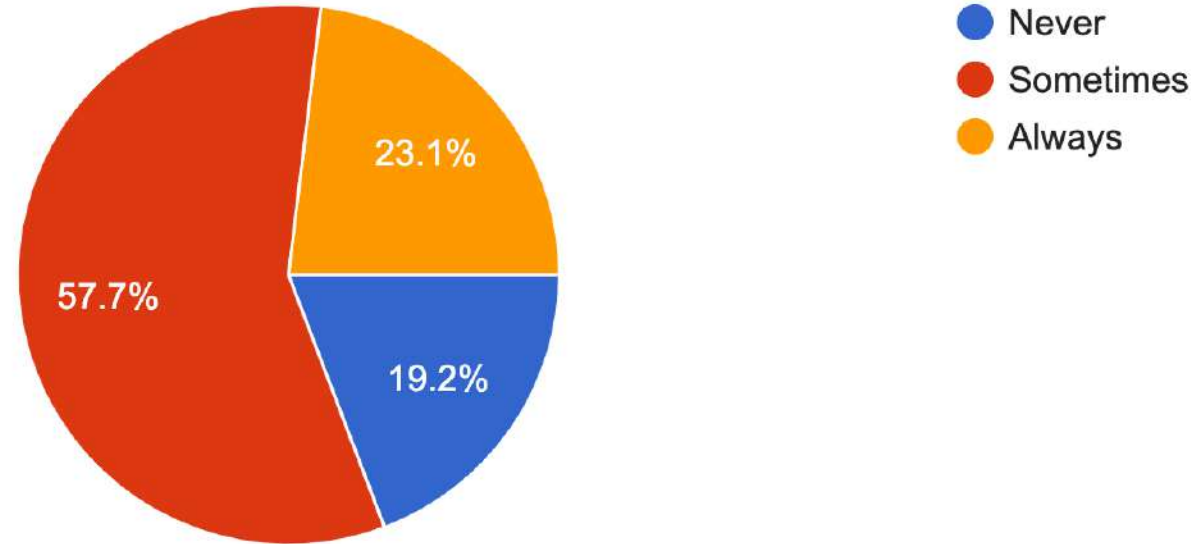
26 responses

# RESULTS



## Do you use machine translation (MT)?

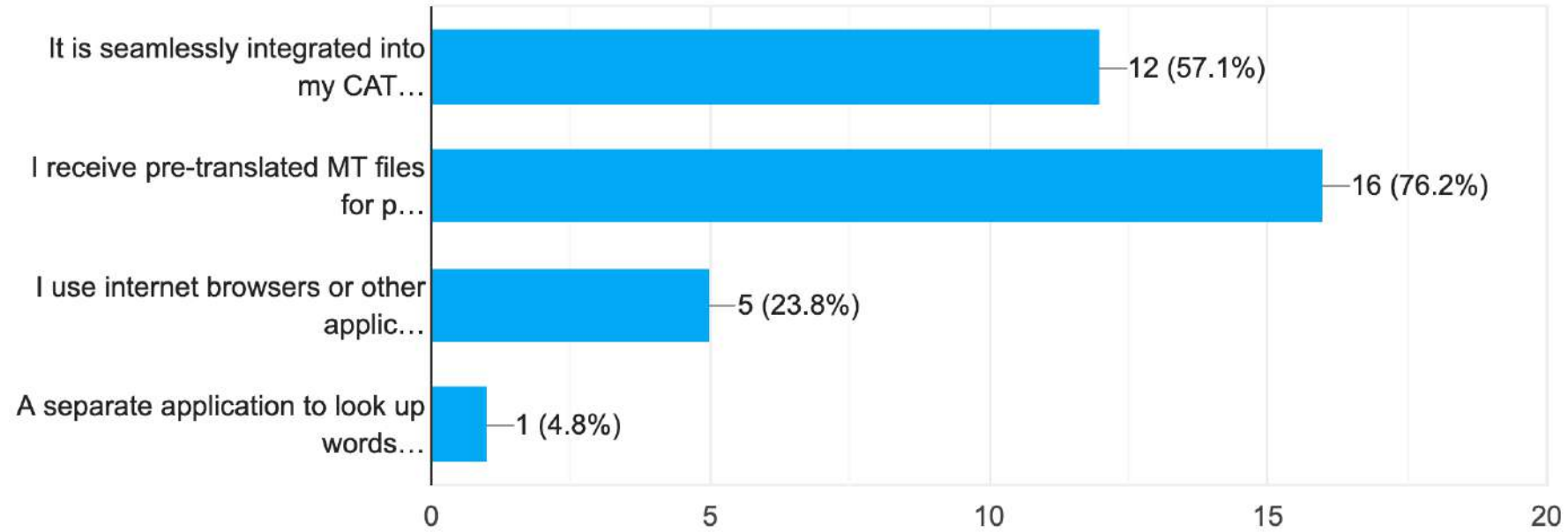
26 responses



**MT USAGE**

## How do you use MT?

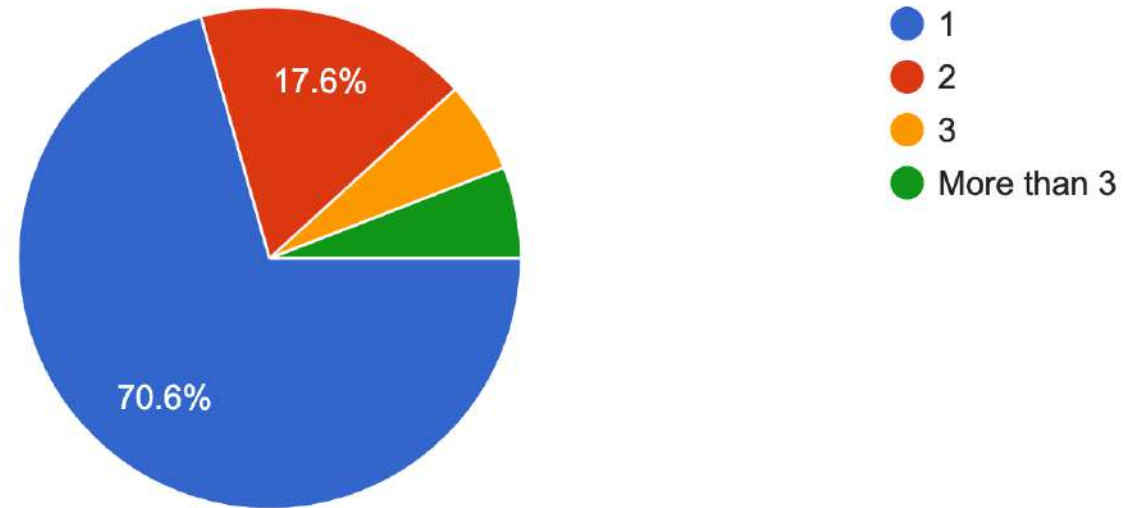
21 responses



**81% USING MT...**

If you have integrated MT into your CAT tool translation workflow, how many MT engines do you use during your translation session?

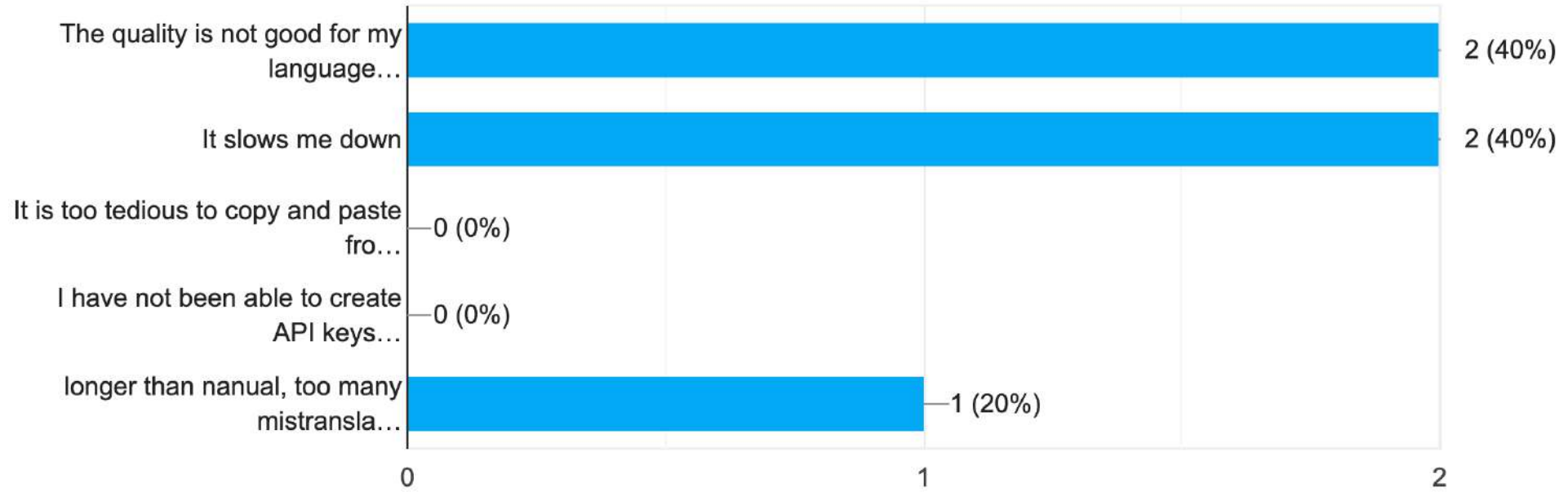
17 responses



**CAT TOOL INTEGRATION**

## Why don't you use MT?

5 responses



**WHY NOT?**

- “I have been using Google and Microsoft for a couple of years and they have improved a lot since then. It was definitely a huge change for my productivity goals!!”
- “I do make use of it, but it saves me only 5% of my time.”
- “Honestly, I once swore to the Gods of translation I would never ever use MT... But the truth is - and it hurts to admit it - MT has been improving tremendously. I would say it's now almost 70% reliable on basic, straightforward documents and with English as source... But yes, all in all, it does save us a lot of time now.”
- “Machine translation has enabled me to accept bigger projects with very short deadlines - so far, so good!”
- “MT is very helpful and makes me more efficient.”

MT GOOD



- “Some agencies use a good MT and it is worth it. Others are so bad that it is more time consuming than translating from scratch.”
- “It's not always very effective. It sometimes saves time on very simple translations but not on very technical or legal documents.”
- “I prefer translating a text from scratch to post-editing MT output generated by a client's or an agency's system... At a certain point, editing out the machine's errors (which tend to multiply the more complex and/or specialized the text is) can become tedious or even irritating because it would have been far more efficient to just have a human translator translate the source text in the first place.”
- “In my pair, to use machine translation compromises the quality of the translation.”

**MT BAD**



**CONCLUSION**

# HYPOTHESES

- Majority using MT? > **Yes**
  - Only 23% always use MT, 68% sometimes
  - 62% of translators perform post-editing jobs
  - Despite having a seamless CAT tool integration, 58% **sometimes** use MT
- Single source? > **Yes**
  - 83% who **always** use MT have a seamless CAT tool integration
  - 73% who have a seamless CAT tool integration use only 1 source of MT

# RISKS



“MT compromises quality”



Belief that correcting MT output is more time consuming or less efficient



“Does not work well with technical or legal documents”



No concerns voiced about confidentiality

## SUGGESTIONS



Provide translators with machine translated texts to convince



Teach translators how to obtain API keys



Encourage translators to get MT post-editing training (shortcut keys)



A man in a dark suit and blue tie is speaking into a microphone. He is looking slightly to his left. The background is a dark, textured wall.

**“WE ARE NOT A  
TRANSLATION COMPANY.  
WE ARE A TECHNOLOGY  
COMPANY.”**

**PHIL SHAWE, OWNER OF TRANSPERFECT  
TRANSLATIONS**

**THANK YOU!**

**JOHN DI RICO**

**@JOHNNYDIRICO**