# DQF BI Bulletin - Q1 2019

Once you start measuring and analyzing the more granular metrics behind your translation workflow, new patterns and correlations emerge and provide you with new perspectives on your production workflows. In this new edition of the DQF Business Intelligence (BI) Bulletin, we step aside from the quarterly trends and instead focus on differences in productivity between translation memory and machine translation.


## Machine Translation as a Part of the Process

The previous edition of the BI Bulletin explored productivity and edit density numbers for each of the segment origin types. The use of translation memory for pre-translation proved to be prevalent in the DQF database. Compared to machine translation assistance, segments translated with translation memory had a shorter edit distance on average and were translated faster.

Machine translation is used in many of the DQF projects, but a relatively small portion of the segments is actually translated with the assistance of machine translation. How does that work? Typically, machine translation and translation memory are given different priorities in the translation process, and therefore interpreting the statistics for both segment origin types will require some context to understand what is going on.

Based on the techniques that are used in the translation workflow, DQF distinguishes between several process types. Users might involve machine translation, translation memory, human translation, or any combination of these techniques in their project. For the projects that have been submitted to DQF since the beginning of 2019, more than 75% included machine translation at some point in the workflow.

However, looking at the segment origin of individual segments, things look completely different. Only 14% of the words in DQF were translated with machine translation as the segment origin, and over 70% had its origin in translation memory.

Although translation memory has been used in the biggest part of translation, relatively little time is spent on translating this content.

A typical workflow that involves machine translation is a hybrid process, in which also translation memory and human translation are used. Human translation can be translating from scratch or editing other human translation. Together this makes up the "MT+PE+TM+HT" process that characterizes the vast majority of DQF projects.

In this hybrid process, translation memory takes the first stab at pre-translating the source text. For each time an exact match for the source text is found in the translation memory, it provides the translator with the available translation of this text. Also, segments with high similarity to segments in the translation memory (high fuzzy matches) are pre-translated the same way. Unlike most exact matches, these are generally expected to require some manual editing. As we will see, this distinction between exact and fuzzy match makes a big difference.

In the hybrid TM/MT process, whenever there are no high fuzzy matches available, machine translation takes over and generates a pretranslation. For that particular segment, DQF flags the segment as 'MT'. Notice that, although the project's process type might be 'MT+PE+TM+HT', it does not follow that each of the techniques is actually used: it is possible that a project with this process type has a translation memory that covers all the projects segments, or none at all.

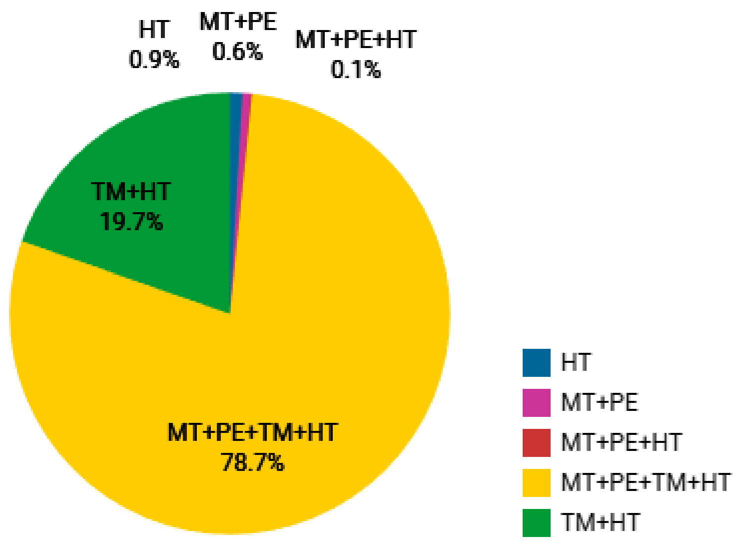


Figure 1: Process used in DQF projects

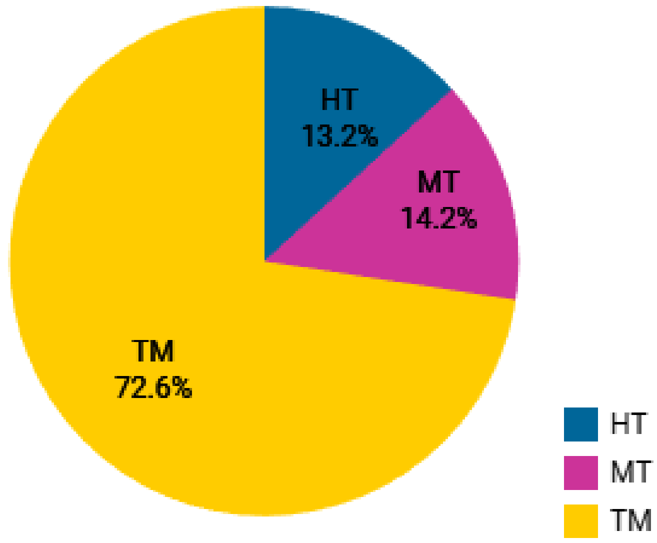


Figure 2: Distribution of words in each segment origin

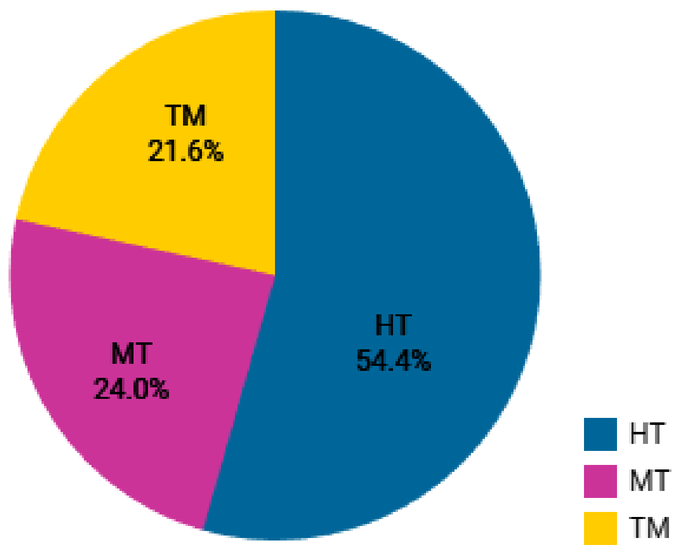


Figure 3: Distribution of time spent on each segment origin

The degree of fuzzy matches can be expressed in a percentage. The exact match has a 100% match rate, and is almost certainly more reliable than any machine translation. However, the lower the match rate from translation memory, the more reliable machine translation becomes in comparison to translation memory.

# Translation Memory and Machine Translation Compared

The DQF dashboard tells a clear story: translation memory as segment origin is more productive, and has a lower edit density than segments pre-translated with machine translation. Certainly, in terms of productivity, the differences are vast.

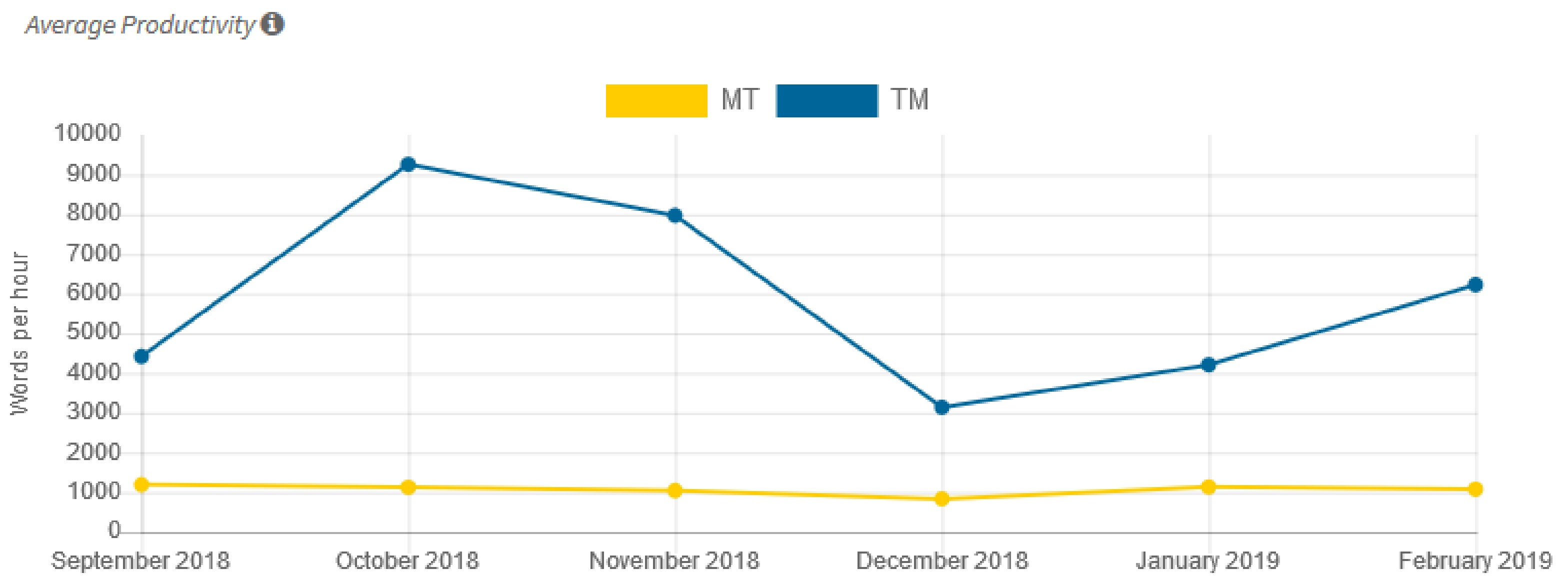


Figure 4: Average productivity in TM and MT

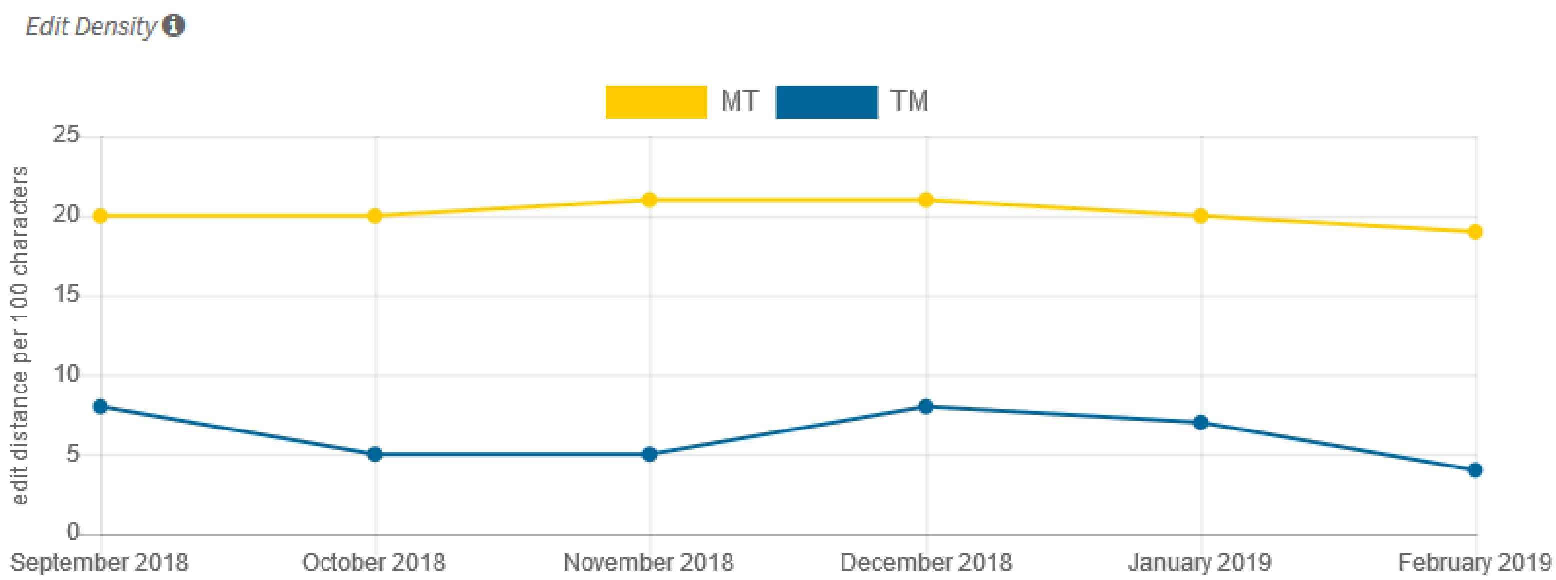


Figure 5: Edit density in TM and MT

But the picture changes dramatically when the segments that did not need any editing during translation are filtered out. Typically, segments from translation memory that do not need any editing are the exact matches.

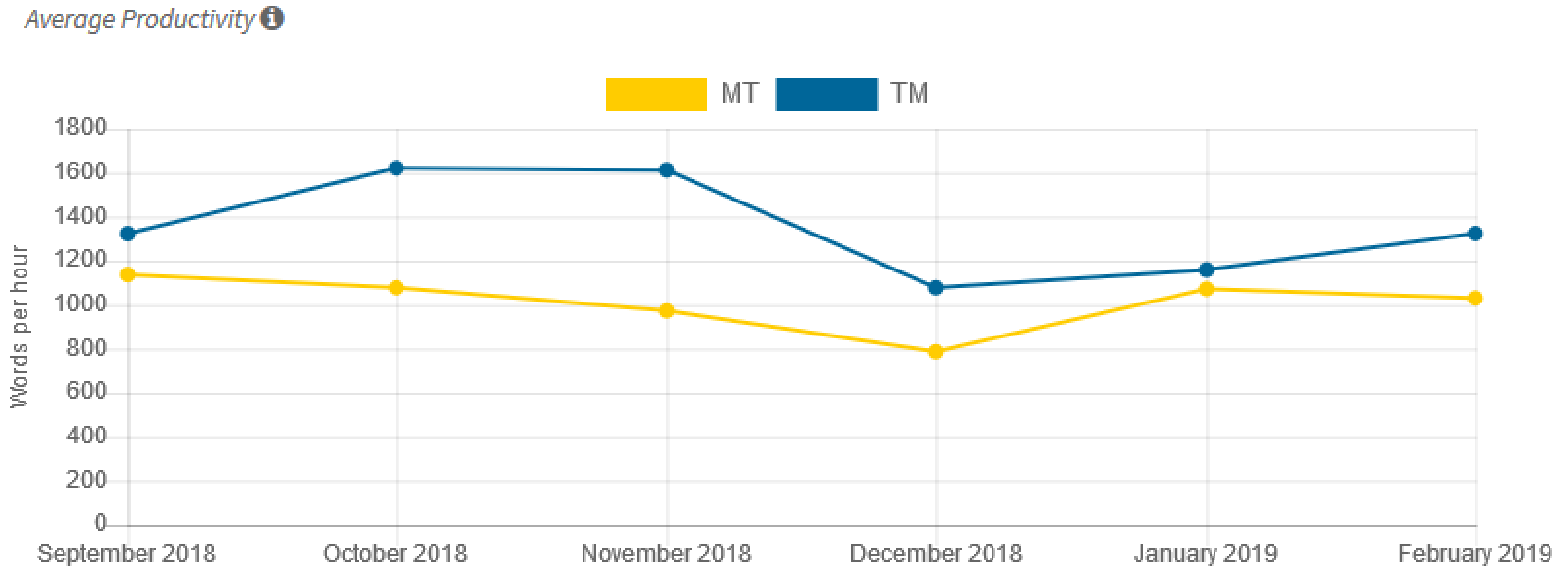


Figure 6: Average productivity for edited segments only

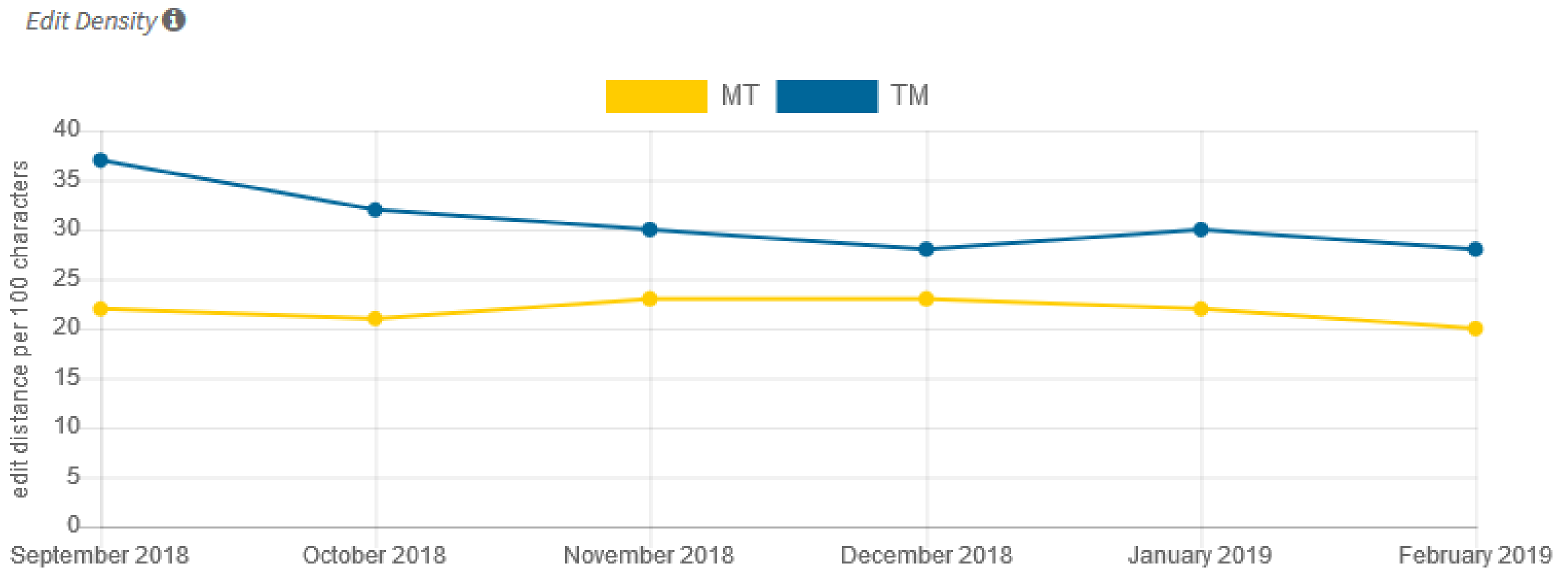


Figure 7: Edit density for edited segments only

What does this show? Segments from translation memory that need editing, have higher edit distance than segments from machine translation. The productivity rates with both segment origin types are now almost the same.

It is surprising, though, that even with a higher edit density, the productivity rate of TM segments is still slightly higher than the productivity rate of MT segments. The most straightforward explanation for this would be that translators edit the target segment more efficiently, if they know the difference between the fuzzy match in the TM and the new source segment. Usually CAT tools give visual clues about the differences between the source segment and TM match. Machine translation cannot give this clue, and translators need to check the complete target text. This might account for the lower productivity in post-edited machine translation, even where the edit distance is lower.

'Fuzzy match' as a category can be quite diverse. It goes from near exact matches, with only a single character difference, up to a point where the match rate is too low to be useful. This usually is a preference setting in a CAT tool, and is traditionally set to a value somewhere around 85%. But does this rule of thumb actually make sense? What is the sweet spot in the hybrid process, what is the ideal cutoff point where translation memory is no longer efficient, and machine translation should take over?

# Where Fuzzy Becomes Too Fuzzy

Let's take a closer look at the fuzzy bands of TM matches. For this investigation, our sample includes the DQF content that was translated after January 1, 2019. In order to be able to compare MT and TM segment categories, we only considered the target languages for which there were both MT and TM as segment origin available. The efficiency of each of the TM match rates is measured by calculating the average time it took to translate 100 characters. It turns out that there is no data available for TM matches below 70%, which means that none of the DQF users use TM when the similarity is below 70%. And that makes sense, as we see that editing matches below 75% already takes much more time than the higher matches. At that particular match rate, the average time spent on translating 100 characters of source text goes from around 40 seconds to more than 100 seconds.

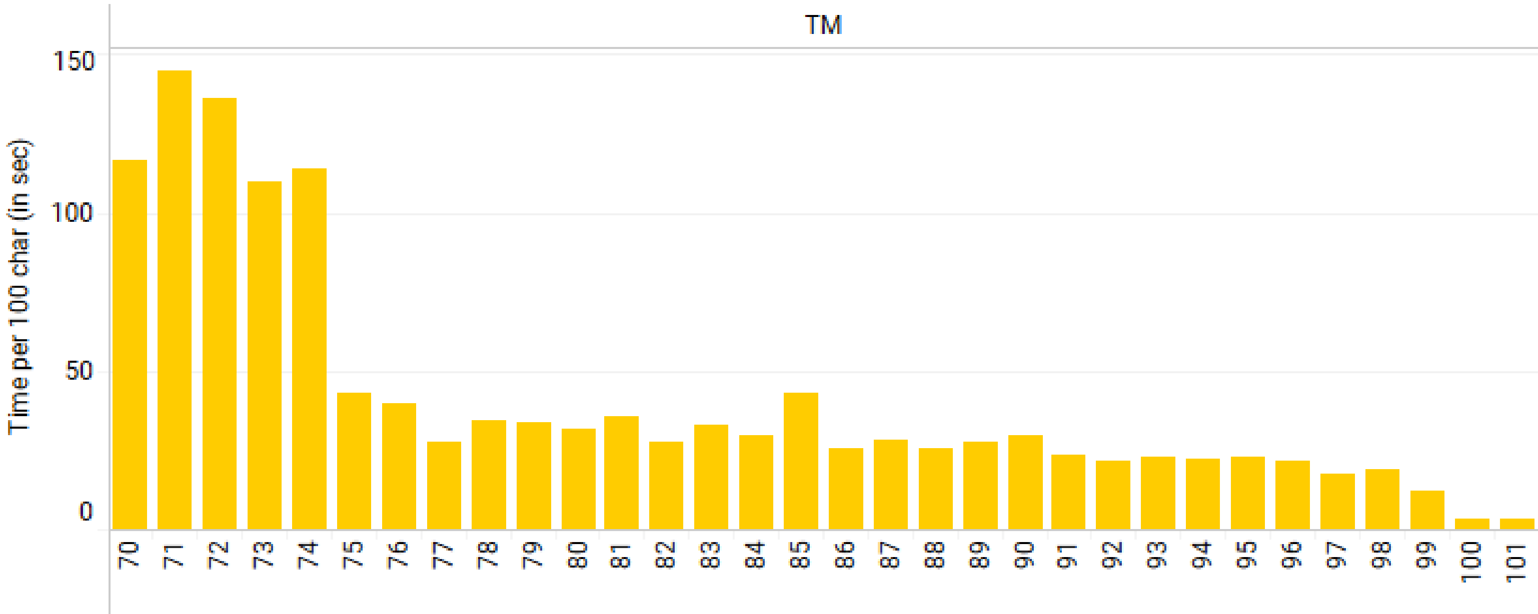


Figure 8: Average time spent on translating 100 characters using TM

The following chart zooms in on the TM matches of 75% and above, and adds MT to it. MT appears to be comparable to TM match rates around 75%.

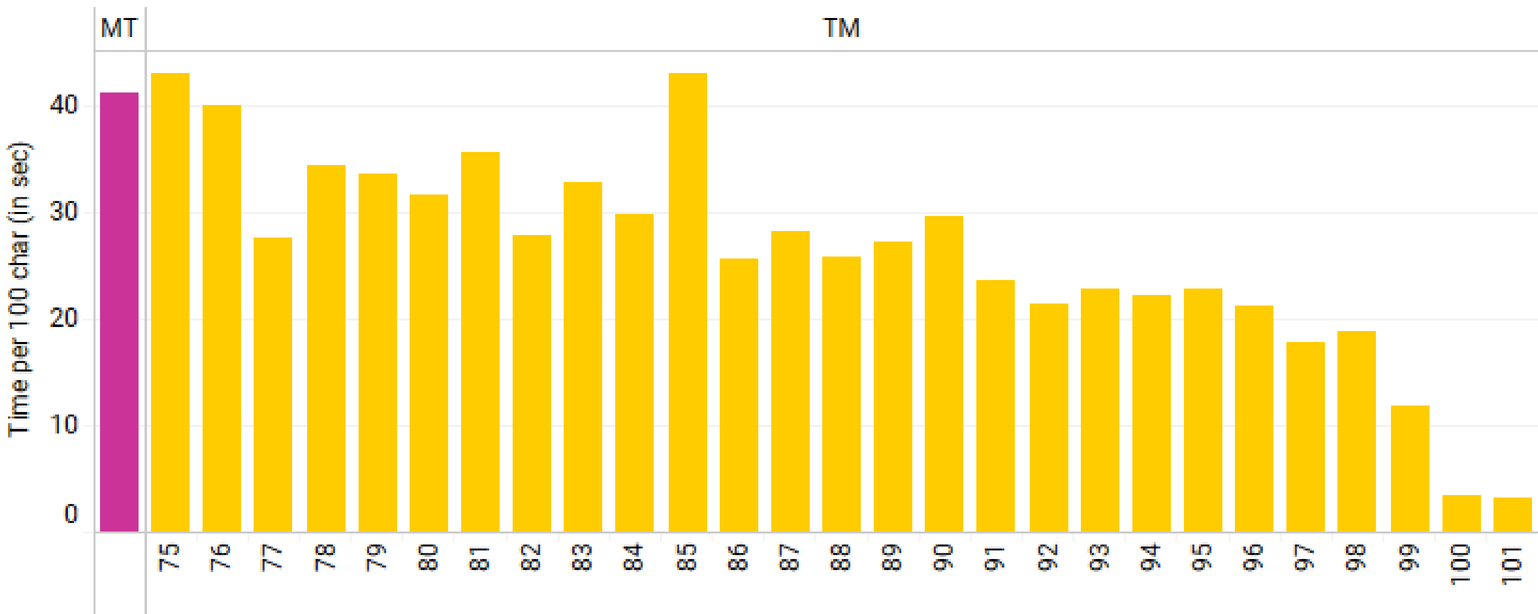


Figure 9: Average time spent on translating 100 characters using TM for higher matches or MT

# Differences Between the Languages Are Substantial

The trends in MT productivity have a quite stable pattern over time, as the trend reports show. But that does not mean that machine translation is equally productive across different languages. There are considerable differences between languages when it comes to the average time that is needed to edit a machine translation for every 100 characters of the source text. Again we took the same sample, and filtered on a few of the bigger target languages that used MT and had English as the source language.

As shown in figure 10, it appears that the MT productivity in the Western-European languages is twice or almost three times as high as in the Asian languages.

Brazilian Portuguese and Spanish being the MT champions in DQF, how do MT and TM compare in these languages? MT here is on par with fuzzy matches between 75 and 85%, and it shows that the sweet spot for switching to machine translation might move up to the higher TM match rates.

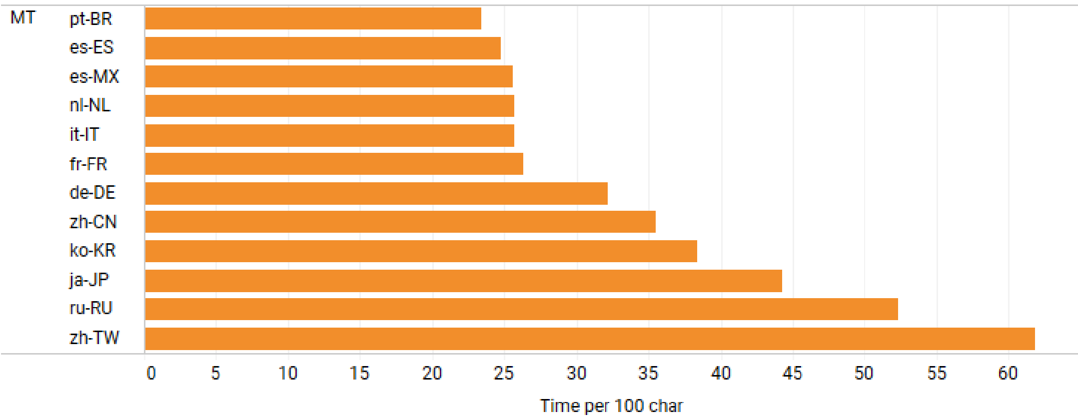


Figure 10: Average time spent on translating 100 characters using MT

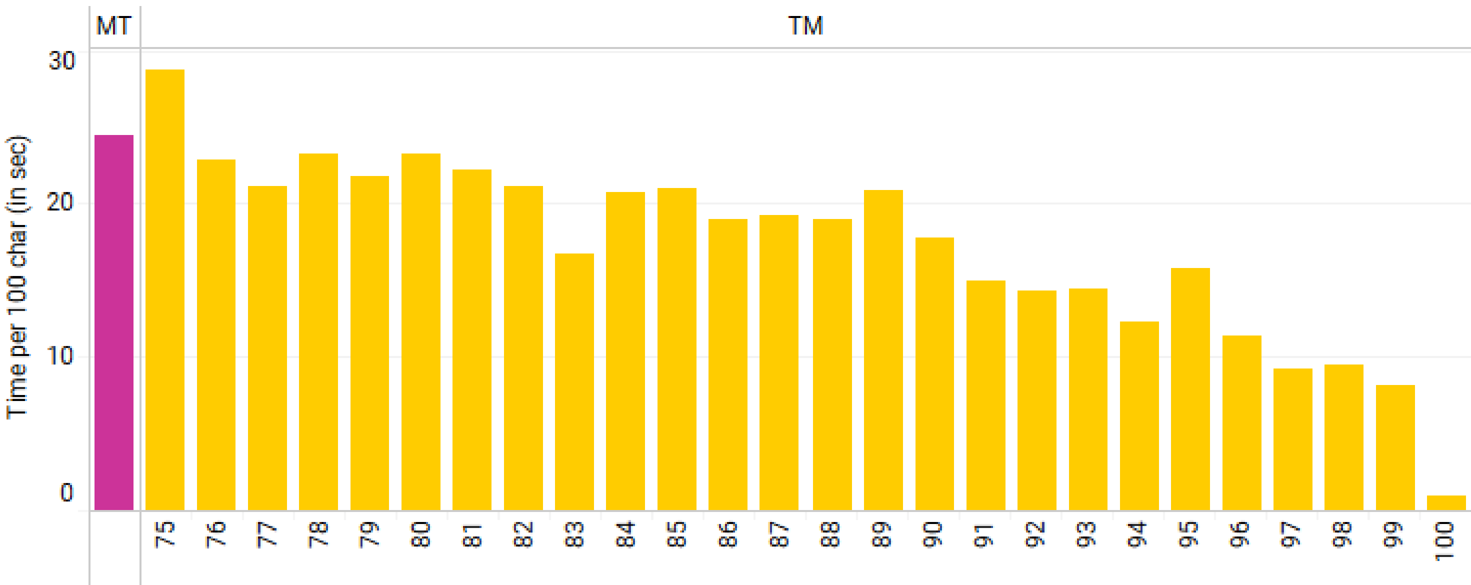


Figure 11: Average time spent on translating 100 characters using TM for higher matches or MT, in Portuguese for Brazil and Spanish for Mexico or Spain

As DQF shows, nothing beats a good translation memory that can provide exact matches for a big portion of the source content. However, the productivity in machine translation comes close and sometimes even catches up with the lower fuzzy matches. It's the right time to revisit the rules of thumb and replace it with proper analysis. You might find some surprises.