10 January 2019

DQF BI Bulletin - Q4 2018

This quarter's averages

13.9 % MT post-edited content

1632 words/hour	960 words/hour	12.3 edits/100 characters	20.4 edits/100 characters
Average Productivity	Average Productivity MT	Average Edit Density	Average Edit Density MT

In this periodical report, TAUS analyzes the most interesting trends in the DQF API data. With more than 150 million words and the many data points that DQF has collected up to now, there is a multitude of conclusions to draw about translation productivity and quality. DQF data answers questions like: what is my average throughput per hour, and how does that compare to the prevalent throughput in the industry? How much of our total volume this past month has gone through machine translation, and how does that compare to the industry average? What has been the effort of my translators working with this specific machine translation engine?

From our interaction with the field, we get a better understanding of the KPIs that quality managers would like to track. Edit density and productivity turn out to be meaningful data points to estimate translation effort. These are the data points that we will be focusing on in this first edition of the new DQF Business Intelligence (BI) Bulletin series, serving as barometer for buyers, providers, and producers of translation around the world.

Translation Memory Still Indispensable

In this edition, we explore the patterns in productivity and edit density from Q4 2018. The main parameters discussed in this bulletin are the three major segment origin types - translation memory, machine translation and human translation, and the two quality levels in DQF - 'high quality' and 'good enough'. These two quality levels indicate the distinction in expected translation quality, and roughly correspond to the distinction between 'light' and 'full' post-editing in the context of machine translation.

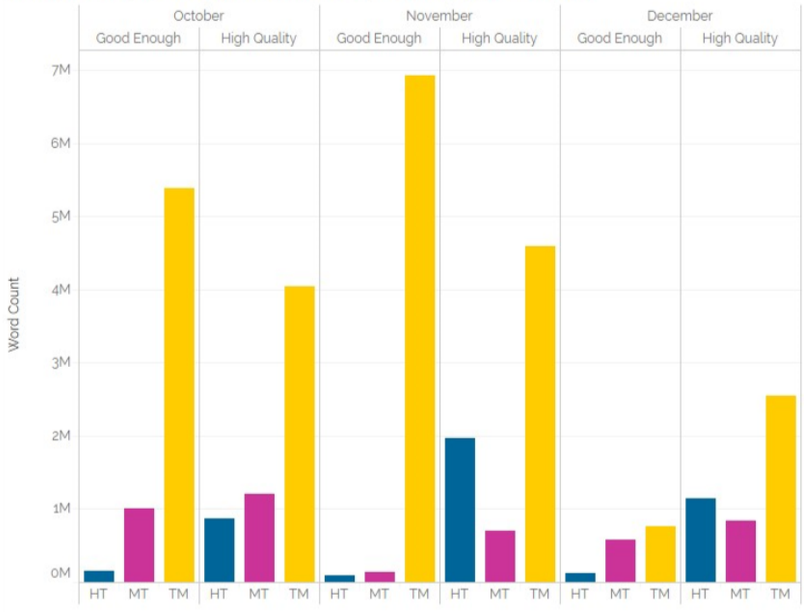


Figure 1: Word count per quality level and segment origin in Q4 2018

If we look at the volumes for each of these categories, we see that translation memory as segment origin is significantly dominating the total numbers for this quarter. Any of the trends in this report are heavily impacted by the statistics from this origin. Both expected quality levels are fairly balanced, with slightly more than half of the translations (52%) falling within the high expected quality level, and the remaining 48% within the lower expected quality level.

Productivity and Edit Density Correlate Strongly

DQF measures productivity as the average amount of words translated per hour. Over the last quarter, the average productivity has decreased from 2,000 to 1,100 words per hour. This was largely caused by the drop in productivity for the projects with lower expected quality level. These projects prove to be much more volatile in terms of productivity rate, and limited series of highly efficient projects can push the productivity rate up for the month. Translation with a high expected level of quality stayed roughly at the same rate of productivity, which is around 1,000 words per hour, with only a slight decrease over the quarter.

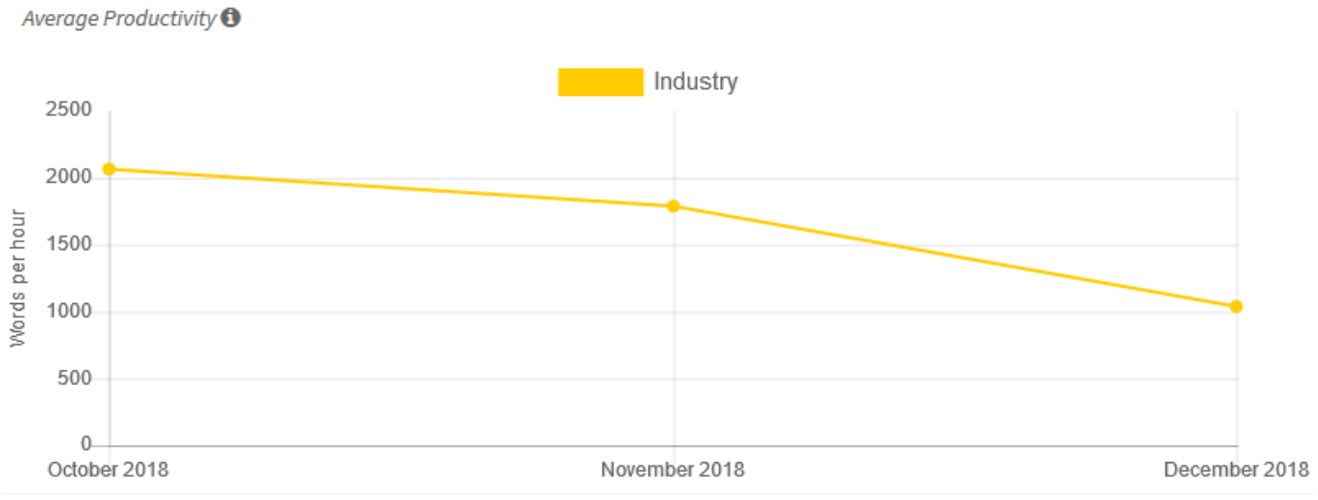


Figure 2: Average productivity rate in Q4 2018

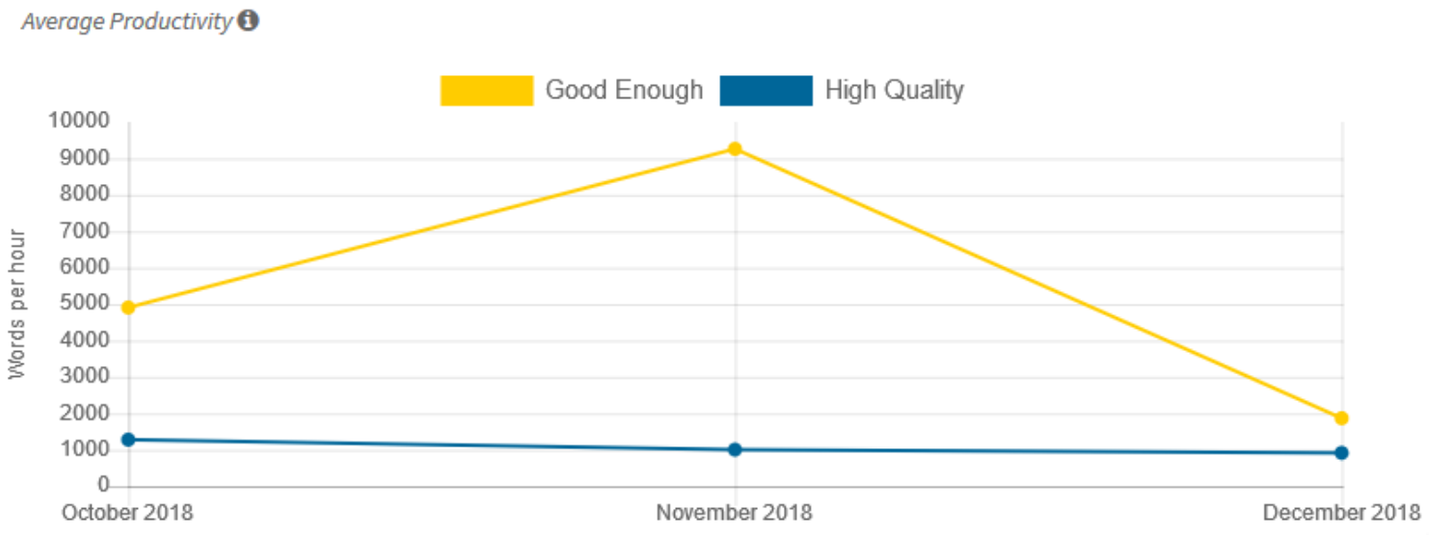


Figure 3: Average productivity rate per quality level in Q4 2018

Evidently, there is a vast difference in productivity rates depending on whether segments got edited by the translator or not. Comparing the two categories, we see that the productivity rate in the edited segments remained largely the same, approximately 600 words per hour, whereas the productivity rate in the non-edited segments shows a trend quite similar to the 'Good Enough' translation projects.

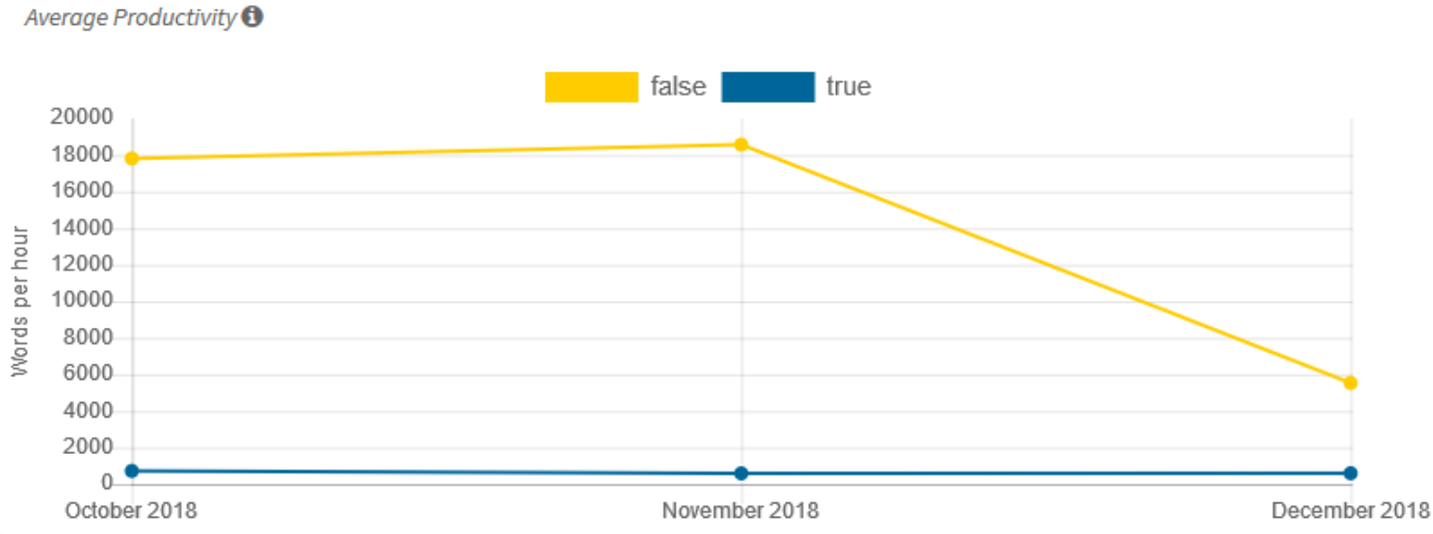


Figure 4: Average productivity rate in edited/non-edited segments in Q4 2018

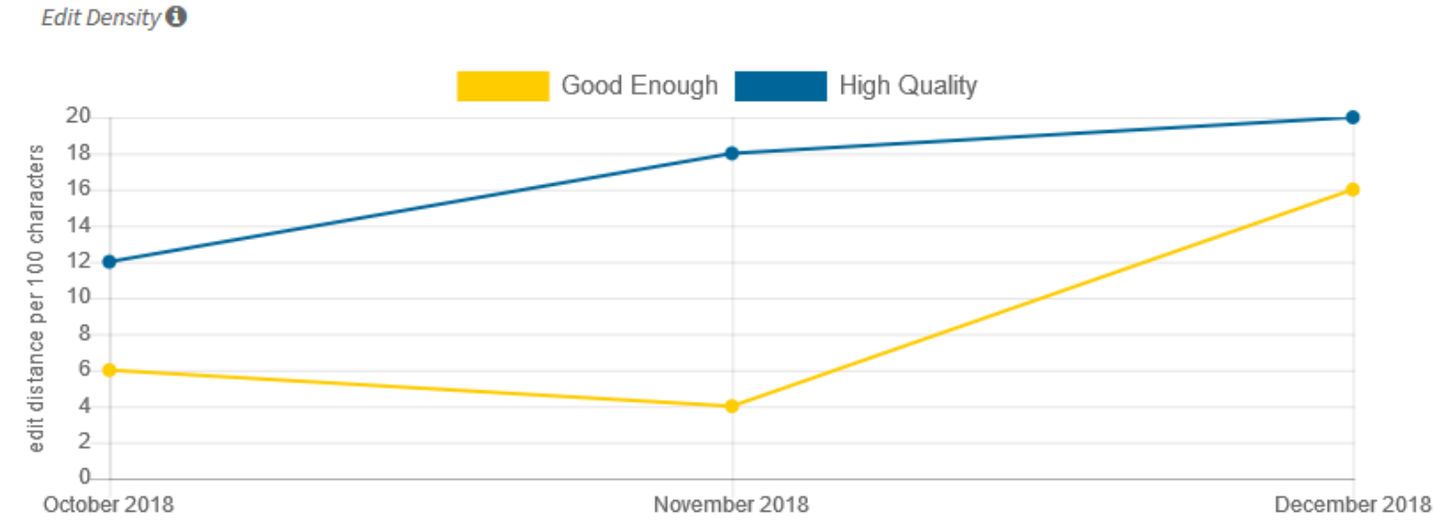


Figure 5: Average edit density per quality level in Q4 2018

With the general productivity slightly dropping, the average edit distance in DQF data went up. The trends here are mirroring one another, because, again, the translation with high expected quality level was more steady in this respect.

A clear connection between productivity and edit density can also be found in the three main segment origin types. With an increase of edit distance, the productivity goes down, which is true for all the segment origins.

The drop in productivity during the quarter can be explained by looking at the translations originating from translation memory. Firstly, translation memory, being the segment origin with the highest productivity, had a smaller portion of the total volume towards the end of the quarter, as can be seen in the processed volumes for the quarter (see figure 1). Secondly, the edit density within this category went up. Both trends amount to the lower productivity rate in December.

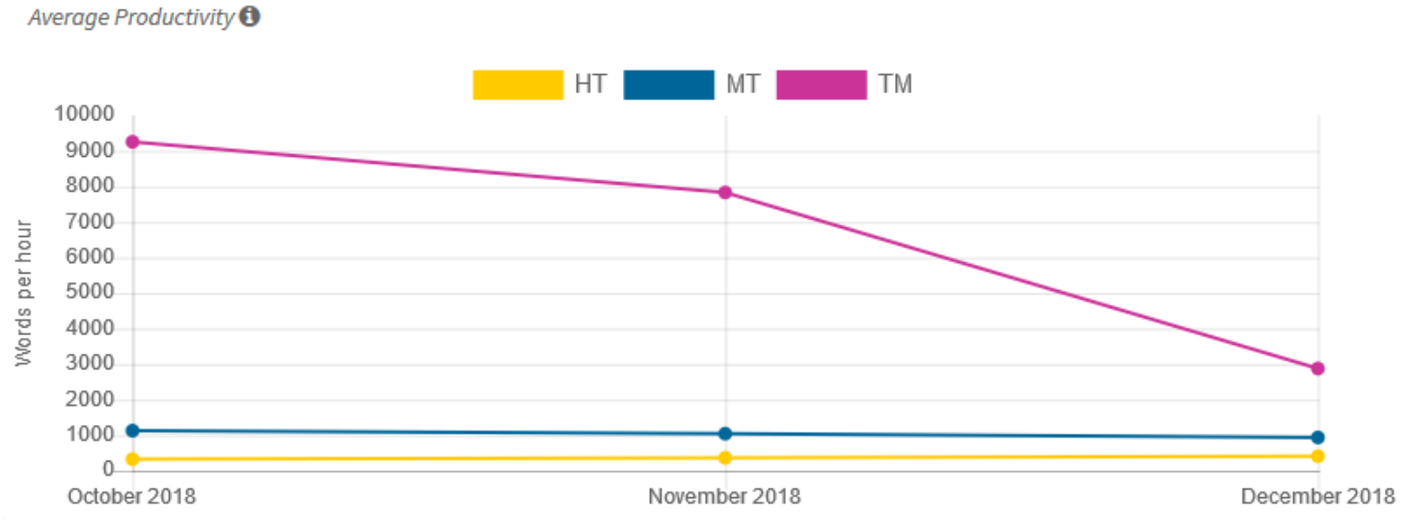


Figure 6: Average productivity rate per segment origin in Q4 2018

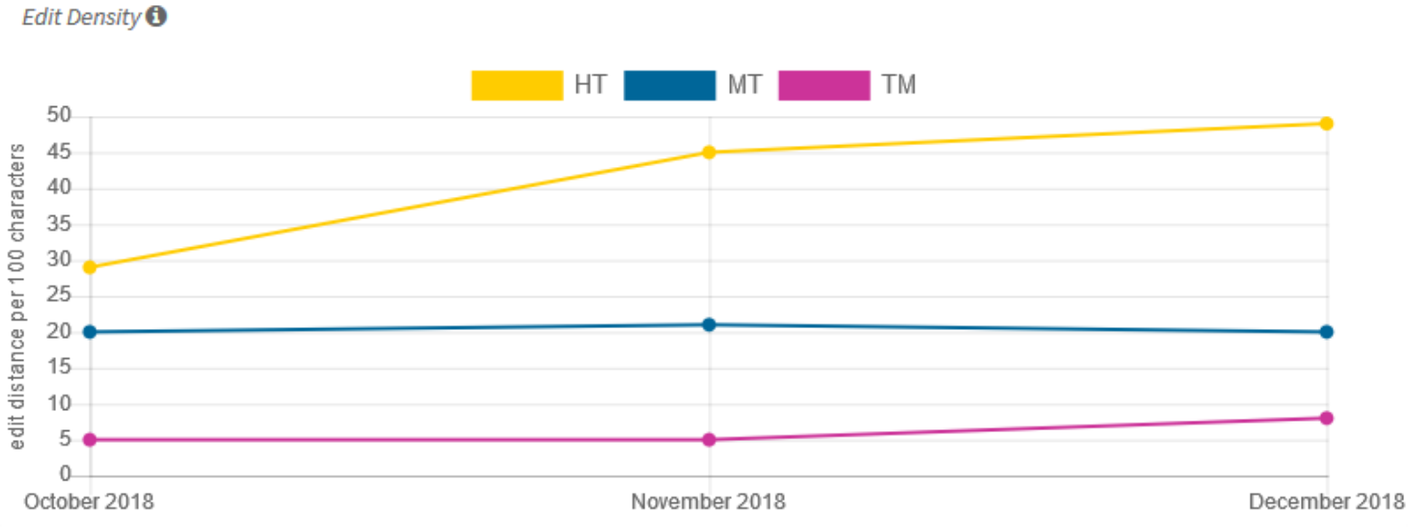


Figure 7: Average edit density per segment origin in Q4 2018

The statistics for this quarter were fairly steady for both October and November, with December deviating from the pattern. Whether this is a seasonal change or not will be explored in the coming editions. Stay tuned for the next BI Bulletin!