1 October 2019

DQF BI Bulletin - Q3 2019



If you are working with different target languages and language vendors in localization, you will quickly see the languages as separate entities that all follow their own paths. Translators for different languages have different price ranges. You get used to the notion of ‘shorter’ and ‘longer’ languages, and you learn the special rules that apply to some languages but not to others. You maintain different language tiers, translation memories, and so on. And bit by bit you get to see the languages as separated and incomparable in many aspects.

With all these different properties for different languages, does that mean that we should refrain from comparing languages altogether? Is there any ground in comparing languages with regard to metrics such as productivity and edit distance? In this BI Bulletin, we will learn the possibilities and impossibilities of cross-language comparison.

Looking for the Common Denominator

Granted that the target languages in your projects are all very different, and have different translation memories and workflows, you might sometimes be wondering if they can be compared according to productivity.

“Ceteris paribus” - all other things being equal - is the leading principle for valid comparisons. In order to measure the influence of one parameter, you need to keep all others equal. DQF as the database for tracking translations can be considered a huge repository of translation tasks. A real live, production environment is not like a laboratory with strictly controllable settings though. But conditions over multiple translations can be similar and comparisons can be made also under similar enough circumstances.

In this bulletin, we consider the DQF data collected in the months of June, July and August 2019. The source language in this period was English for 97% of the translations (93% American English, and 4% British English). For this reason, this analysis focuses only on English as a source language.

The DQF database has 85 different target languages in total. The main languages among these are in the chart (Figure 1) on the right.

DQF uses a few basic metadata to describe translation projects. It is these metadata that allow us to create fair conditions for our comparison. Let's have a look at a few of these metadata (Figure 2-4) to determine which categories are the most frequently used, and are therefore suitable for our research sample. DQF distinguishes between the sectors of industry, the content type, and the expected quality level.

For each segment translation in DQF, the segment origin can be determined. This indicates the source of the pretranslation. This can be either machine translation (MT), translation memory (TM) or human translation (in which case there was no pretranslation).

As can be concluded from the numbers in the pie charts, in the given period DQF got most of its data from the cloud services and computer software sector, from projects aimed at a high-quality level, and from both user manual and marketing content types.

Languages might not be evenly distributed among each of these metadata. The metadata can be combined with the volumes in each language to see which intersection gives a large enough body to compare data.

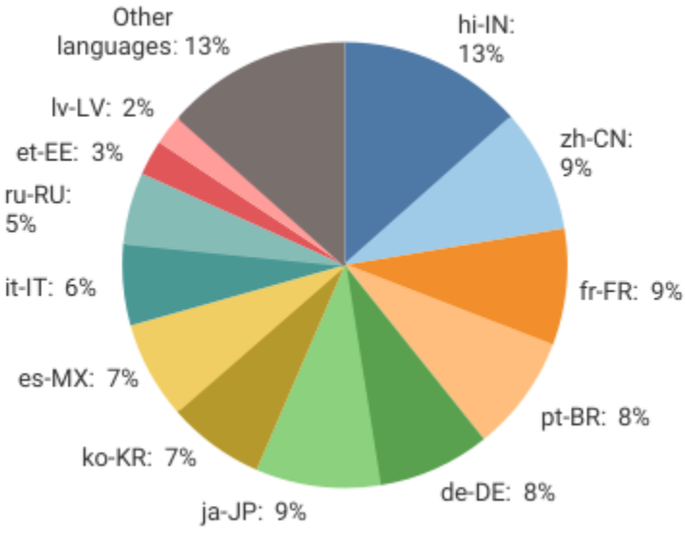


Figure 1: Most occurring target languages in DQF

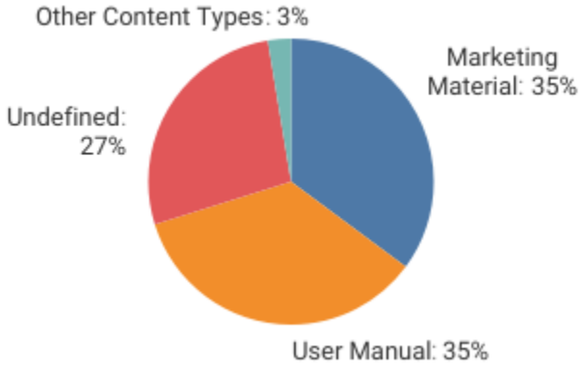


Figure 2: Most occurring content types in DQF

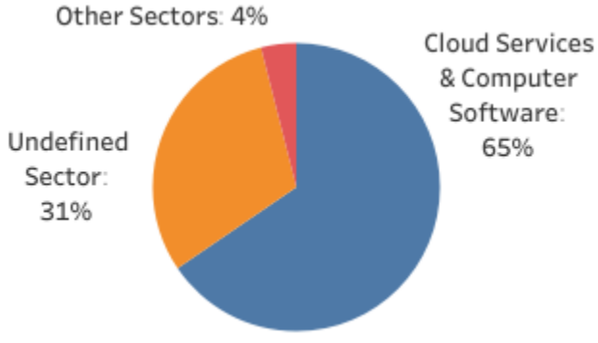


Figure 3: Most occurring sectors in DQF

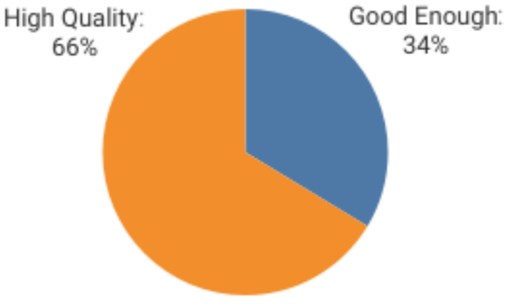


Figure 4: Volume of translated words in both quality levels in DQF

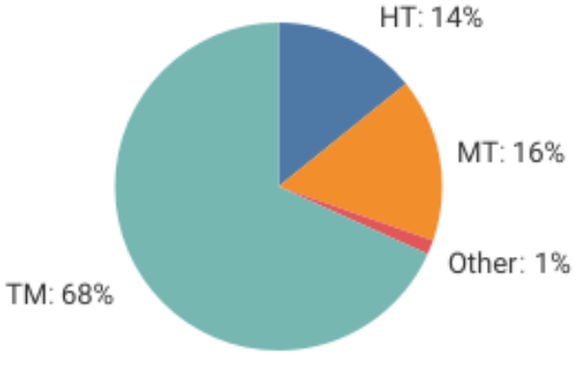


Figure 5: Volume of translated words per segment origin in DQF



Figure 6: Target languages with the largest volume in June, July and August 2019

From this data the following considerations were made to select the research sample:

- High Quality translations are preferred over Good Enough since we assume that translators will be better motivated to edit translations as they see fit without any further constraint.
- For sector, the categories ‘undefined’ and ‘other’ do not offer consistent data, leaving only Cloud services and computer software as categories for the sector.
- Content type ‘user manual’ is lacking in the high-quality category, leaving only the marketing content type.
- The target language with the biggest volume in DQF, Hindi, doesn’t have enough translations in the required content type and quality level to be included in the sample. However, the target languages Chinese, French, Brazilian Portuguese, German, Japanese, Korean, Mexican Spanish, Italian and Russian are well represented in these categories.
- Segments with translation memory as segment origin are internally too diverse and should be excluded from the sample. For example, a pretranslation with a 100% match from translation memory is hardly comparable to a pretranslation with an 85% match.

These considerations limits the sample to high quality translations with marketing as content type, and cloud services or computer software as a sector, with machine translation as segment origin.

Where Productivity Meets Edit Density

With all the filters ready, we can make comparisons across the target languages.The sample we have built up so far has a wide variety in productivity, ranging from 1740 words per hour for translations into Brazilian Portuguese or Italian, to 828 words per hour for translations into Japanese.

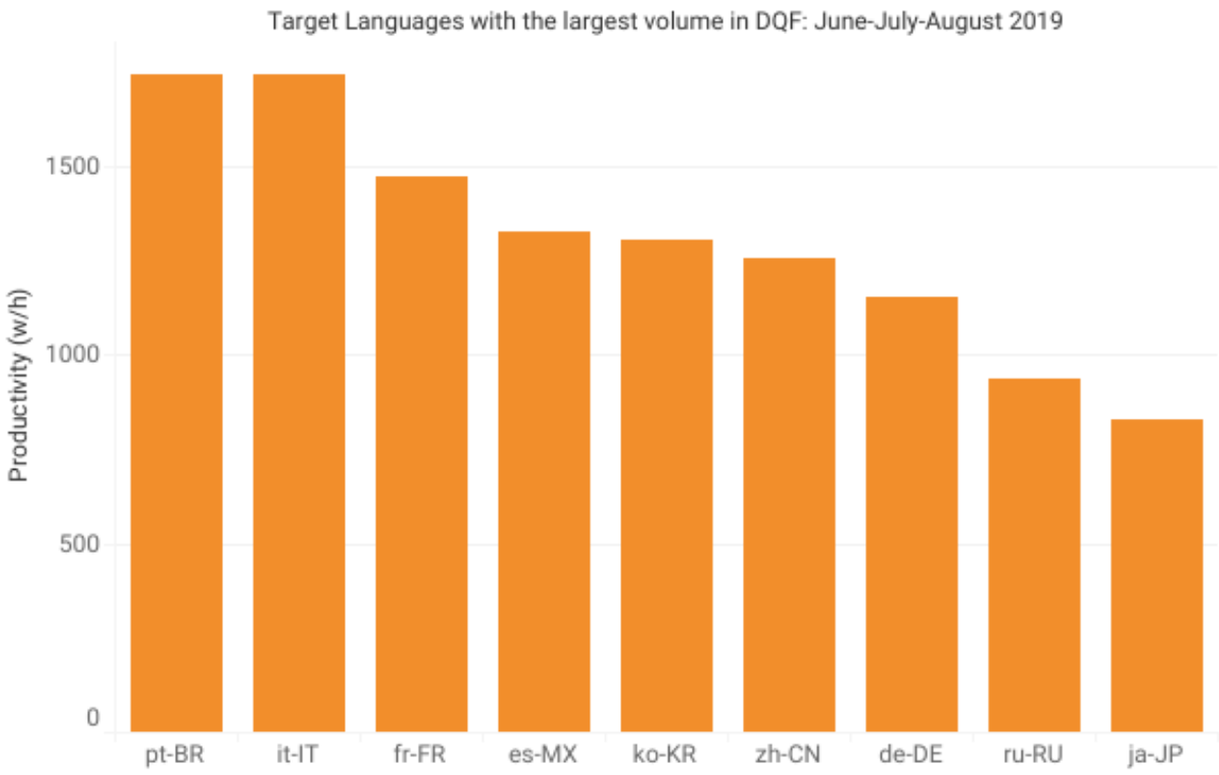


Figure 7: Productivity for the target languages with the largest volume in DQF, June-July-August 2019

The immediate explanation for the variety in productivity is that for each of these target languages the quality of the machine translation output differs and impacts the speed of translation. The machine translation output can be analyzed by looking at the relative number of edits that are needed in these languages. DQF uses 'edit density' for this, which is calculated by averaging the number of edits per 100 characters in the target language. Good quality MT suggestions most likely result in a lower edit density during translation.

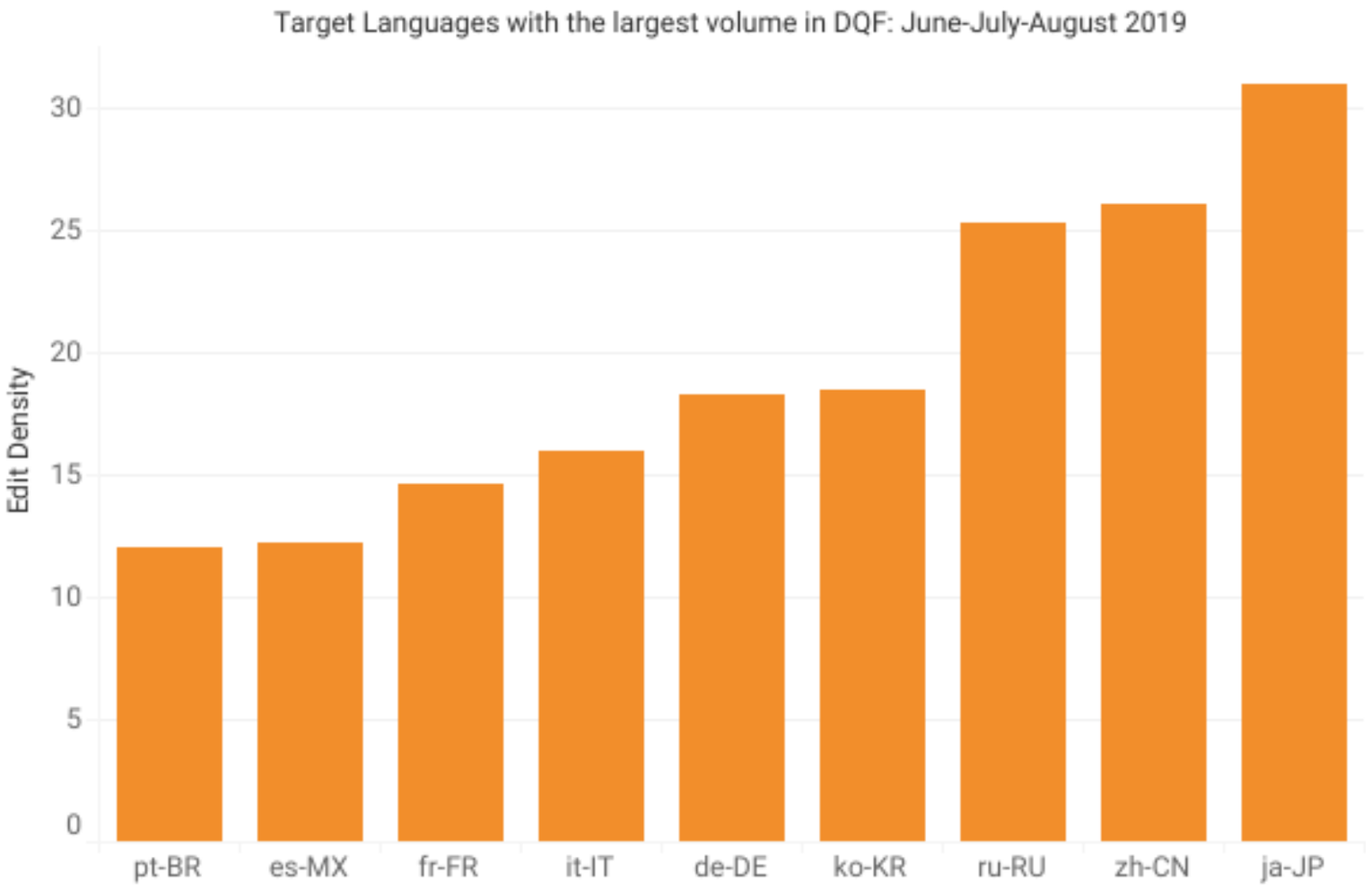


Figure 8: Edit density for the target languages with the largest volume in DQF, June-July-August 2019

Calculating edit density for the target languages shows that, as expected, in many cases high productivity goes hand-in-hand with a low edit density. It is the main trend, and it is true for Brazilian Portuguese and for Japanese at either end of the continuum. But comparing the languages, we can also see some differences. For example, Italian has high productivity, nearly the same as Brazilian Portuguese, but the edit density is much higher. The opposite is true for Mexican Spanish, where despite the low edit density, the productivity is not as high. A scatter plot can visualize the relations more clearly, as it shows how researched entities score on two dimensions. The scatter plot below shows for each language both the edit density and productivity.

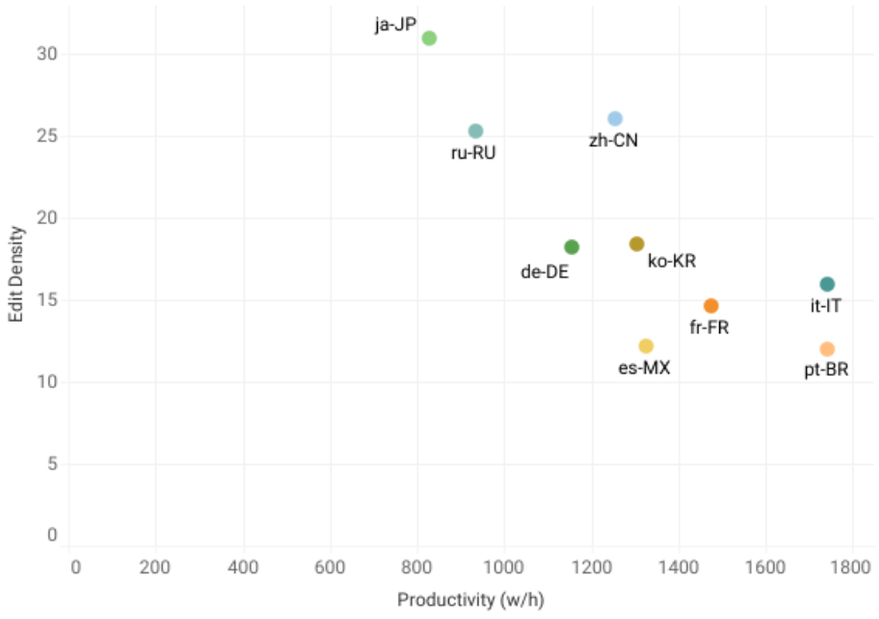


Figure 9: Productivity and edit density for the target languages with the largest volume in DQF, June-July-August 2019

The general trend in this scatter plot is clearly going from high edit density with low productivity to low edit density with high productivity. However, the chart also shows that some languages relatively similar to another language on one dimension might differ on the other dimension. For example, with a higher edit density, Italian equals Brazilian Portuguese in productivity, whereas Brazilian Portuguese has higher productivity than Mexican Spanish, despite their edit densities being similar.

Though cross-language analysis can take us quite far, some caution is needed here: productivity measures the number of translated words in the common source language, but edit density is measured in the target language. This introduces language-dependent features to the comparison. In this sense, it is interesting to look at language-specific characteristics, such as the relative language length. This can be expressed in the average number of characters per segment, as shown below:

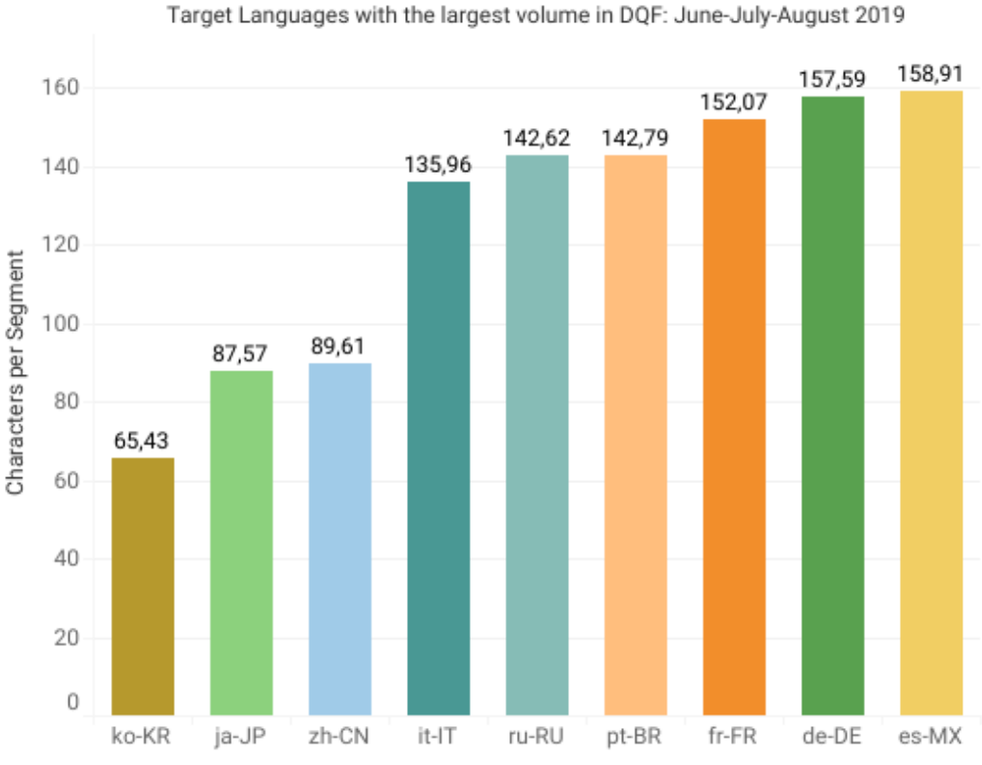


Figure 10: Average segment length per language

For the pairs of languages that have similar edit densities (Russian and Chinese, German and Korean, French and Italian, Mexican Spanish and Brazilian Portuguese), it is consistently the longest language that has the lowest productivity.

What exactly is the influence of the length of the target language on the productivity cannot be determined, based on this short research? Nevertheless, it makes good sense to include properties as the average segment length in the cross-language analysis.

As this short investigation shows, cross-language comparison is possible, and can give some useful insights in each of the separate languages. To do this, limit the research sample to a strict set of conditions using the DQF metadata, to make sure that you create an even playfield, and rule out those influences that can distort the results. Make sure that you stay within the same content type, sector and segment origin. If some of the calculations use language-specific features extra care is needed to avoid drawing conclusions too rigidly.

DQF offers the toolbox to measure the cross-language trends, but also gives the opportunity to further drill down to the specifics of each language to make sensible decisions based on real live numbers. Schedule a demo to find out more.

Request a DQF demo to find out more.

REQUEST DEMO NOW