

Machine Translation in the Financial Services Industry: a Case Study

Mara Nunziatini

Welocalize Italy S.r.l.

Via Alserio 22, Milan, Italy

mara.nunziatini@welocalize.com

Abstract

The use of Machine Translation is spreading quickly in the translation industry. While its implementation is smooth in some contexts, in the regulated services industry it certainly seems trickier. In particular, the financial services industry can be considered a less conventional scenario within which to implement MT. This paper explains how MT was successfully implemented in the workflow of a translation company specialized in financial services, and how freelance translators got positively involved in the process.

1 Introduction

Welocalize Italy S.r.l. is Welocalize's Italian headquarter, based in Milan. This hub used to be a translation company on its own and was recently merged into Welocalize to become its FSI-specialized translation hub. The greater part of its business, since its foundation, has always been focused on the translation of financial, tax and legal documents. In order to stay in step with competitors and trends in the translation industry, and to offer a wider range of services and more flexibility to its customers, the company started thinking of implementing Machine Translation in its workflow. In this paper we will describe how we implemented MT in this regulated sector.

2 Description of the Company

2.1 Client Base

The company's client base boasts a relevant number of faithful, long-time customers which mainly includes large Italian enterprises, SMEs

and multi-national auditing companies (or their Italian subsidiaries), but also banks and lawyers.

Every year, for marketing, tax or legal purposes, these customers need to translate financial texts (mainly Financial Statements and Transfer Pricing documentation) and/or legal documents and, after years of cooperation, many clients have become familiar with the company's Project Managers and salespeople. However, in order to retain key customers, find new clients and keep up with competition, the company started facing the need of providing lower prices and shorter turnaround times, while still delivering top quality – a fundamental aspect of the FSI industry. In the meantime, the global translation industry started talking about neural MT (Castilho et al., 2017), and all the major LSPs were already implementing MT in their workflow, therefore offering more competitive prices and shorter turnaround times. Similarly, the LSP we are part of has been using MT for many years, and was already implementing neural MT (Schmidt and Marg, 2018).

2.2 Description of the Business

Our company's business concentrates in one peak season which approximately starts in March and ends in June, this is the time when Italian enterprises (or foreign enterprises with an Italian subsidiary) listed on the stock market publish their Financial Statements in Italian and in English, so as to reach a bigger number of stakeholders. Another busy season corresponds with the publication of half-yearly financial statements (end of summer till October, approximately).

Another relevant part of the business, but with reduced volumes compared to Financial Statements, regards Transfer Pricing documentation. For this type of content, there is not really a peak season – rather, these translation

© 2019 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CCBY-ND.

requests come in continuously, in a more or less regular fashion. Transfer Pricing documents are aimed at proving that the prices of goods and services which are exchanged among subsidiaries, affiliates or controlled companies are in line with the arm's length principles¹. These are usually drafted in English in the case of multi-national companies and need to be translated into Italian for tax purposes.

The greater part of Financial Statements is translated from Italian into English, while the greater part of Transfer Pricing documentation is translated from English into Italian.

2.3 Technology Resources

As for technology resources, the company was not advanced before the merger. Only recently has this business started using CAT tools as part of the standard workflow. In fact, this hub used to implement a traditional translation and proofreading workflow with automation of processes being non-existent. Documents to be translated were sent by email to the translator, who was asked to deliver a final file in the same format and layout as the original. Besides, the Translation Management System in use offered basic functionalities only. For this reason, projects were handled without the file having been uploaded to any CAT tool, and management of resources like TMs and glossaries was completely manual, time-consuming and not so efficient.

After the merger, the use of CAT tools and bilingual files started to be implemented in the workflow, TM management became more efficient, while in the translation industry MT was a topic more and more discussed.

Thanks to implementing a new TMS and creating centralized translation memories in a more structured way, we managed to build up significant and good-quality TMs for the main subjects translated (Financial Statements, Transfer Pricing, Non-Financial Statements, etc.). These memories were very helpful when exploring the use of MT for our group.

2.4 External Vendors

Vendor database is not very big and mainly comprises reliable English and Italian native speakers who have been working for the company

for several years already. These vendors specialize in translating financial documents and have a high productivity in terms of words translated per day, thanks to their multi-year experience and personal linguistic resources.

However, the vendor base is so small that Project Managers end up working always with the same translators, who often get fully booked, especially during peak season. The average freelancer is highly experienced in the subject matter and, due to our own processes, some of them were only recently introduced to the use of CAT tools and other automation aids. The greater part of them had never heard anything about MT.

The small vendor base started to be a blocker for the growth of the company. It is more and more difficult to find financial translators with enough experience who are reliable and affordable, so production really depends on this small base's availability. Furthermore, while the trend in the industry overall is to cut costs, these experienced freelancers tend to increase their rates.

3. Need to Implement MT and Challenges to Implementation

The possibility of offering MT as part of the services has begun to look attractive and indeed necessary, but is it possible to implement MT in the FSI?

3.1 The Importance of Quality

Financial translation requires great attention to details as even a small mistake can lead to a major problem. Financial translation requires expertise and experience, as the importance of integrity and accuracy of information in financial documents cannot be underestimated.

3.2 Terminology

Translating terminology, and doing so consistently, is a major challenge in the financial world. It is fundamental to ensure consistency and comparability between documents of the same company related to different periods (i.e. to compare quarterly and annual reports). Financial terms can be intricate and represent a challenge for translators who do not have understanding of or experience in financial translation. Understanding concepts in their context is very important in

¹ OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations, July 2017.

financial translation – and we know this is one of the weak points of MT. Just to mention a couple of tricky examples, “ammortizzare” in Italian is translated “amortize” if we are talking of intangible assets, while is translated “depreciate” if we are talking about tangible assets. Another tricky one is “periodo” which is translated “year” in yearly Financial Statements but “period” in half-yearly Financial Statements. Terminology must also be compliant with IFRS² (International Financial Reporting Standards), i.e. a set of accounting standards developed by the IASB (International Accounting Standards Board).

Besides, date format and currency format may vary a lot depending on customer’s preference (as at 31 December 2018, as of December 31, 2018...) just like currency format (EUR, €, euro, Euro).

3.3 Numbers Localization

In financial documents, numbers matter greatly. Besides, when doing financial translation, numbers must be localized (Italian and English use different decimal and thousand separators). An error in the positioning of a comma, an excess digit or omission of a digit mean thousands in monetary losses.

3.4 Timeliness

During peak season, the business becomes especially fast-paced and constrained by time, lots of requests come in every day that add on to the already booked translations. Translations need to be delivered in a very short time as it is very important that these documents do not miss any deadline. However, in most of the cases, the greater part of trusted translators and reviewers are already fully-booked. For this reason, salespeople cannot give clients the translation they want in the time they need and are left with nothing better to offer than a longer turnaround time or a lower quality.

3.5 Confidentiality

Financial documents need to be secure since they disclose private company information. They must remain private and handing them over to third parties poses great risks. It is extremely important that the company uses reliable tools, since the LSP must ensure that no data are shared externally.

4 Description of Engine Selection and MT Implementation Process

After having identified all the possible requirements and challenges, we decided to start the engine selection process with the help of the company’s Machine Translation team.

Our ideal candidate was a state-of-the-art, customizable engine which is compatible with the CAT tools used internally. Besides, in order to be cost-effective for production, it must deliver good-quality output. Last but not least, the engine must by no means represent a risk for data privacy, and its price had to be in line with the company’s budget. In 2018, when the implementation process began, “state-of-the-art” meant “neural”.

The potential candidates identified were 3:

- Option 1: a generic financial neural machine translation engine;
- Option 2: a generic non-customizable neural machine translation system;
- Option 3: a customizable neural machine translation provider, which allowed us to create two engines (one to translate Transfer Pricing documentation from English into Italian, and another to translate Financial Statements from Italian into English).

Option 1 and 2 were the first options to come in, while Option 3 was identified only at a later stage and trained with our TMs. All the three Options are neural engines, but at a first glance we would think Option 3 would suit us best as it is customizable. However, the most important criteria to choose the best engine was the quality of the output, so we proceeded to test the quality of each engine’s raw output.

4.1 Testing Option 1 vs. Option 2

The quality test was run on a 2500-word sample from a Financial Statements which was translated from English to Italian with both options. Quality check consisted of a full post-editing of both raw outputs by two native speaker in-house post-editors specialized in financial translation. Quality was evaluated by comparing the amount and type of changes, and the time linguists spent to fix them was calculated. As for the types of mistakes, we noticed that certain issues appeared both in Option 1 and Option 2’s output. The linguists flagged

² <https://www.ifrs.org/>

more or less the same amount of grammar issues, untranslated content, mistranslations and inconsistent terminology in both outputs. Option 1's output showed a bigger amount of formatting issues and omissions, while Option 2's output, being Option 2 a generic engine, showed a bigger amount of key terminology issues.

To sum up, both engines proved to have pros and cons, and we decided to think of what kind of mistakes were quicker and easier to spot and fix. Formatting can be fixed pretty quickly, and terminology can also be fixed easily by connecting a glossary to the project, while omissions are the trickiest issues. For this reason, and also because Option 1 was not compatible with the CAT tools used internally, we decided that Option 2 would be a better candidate.

4.2 Testing Option 2 vs. Option 3

Then, Option 3 was also proposed by the company's Machine Translation team and new tests were carried out to evaluate the quality of Option 3's output compared to the two original options. The second test phase was divided into 2 steps: automatic scoring and human evaluation.

For automatic scoring, we use a proprietary tool that outputs a number of industry-standard automatic metrics, such as BLEU, GTM, Meteor, NIST, PE Distance, TER (TAUS, 2012). We typically run this tool on two sets of input: source + MT vs. human reference from a TM (during engine building), as well as source + MT vs. human post-edited reference (during pilot and production). The table below shows the results from scenario 1, i.e. the human reference was not specifically created by translators performing PE on the MT output.

LP	MT Engine	BLEU	NIST	METEOR	GTM	Avg. PE
EN>IT	Option 1	22.68	5.55	40.56	54.07	43.30%
EN>IT	Option 2	29.79	6.64	47.78	60.63	38.18%
EN>IT	Option 3	39.12	7.62	56.15	66.99	31.62%
IT>EN	Option 1	35.42	7.14	34.42	66.96	42.90%
IT>EN	Option 2	38.78	7.33	38.11	70	39.93%
IT>EN	Option 3	36.21	6.98	35.07	67.18	42.38%

Table 1. Results of automatic scoring.

As showed in the table, Option 1 obtained the worst score for both language pairs, Option 2 obtained the best score for Italian into English and Option 3 obtained the best score for English into Italian.

After human evaluation, Option 1 was excluded again, and Option 2 was also excluded

since the quality did not prove to be significantly better than Option 3 for Italian into English. Option 3 was the preferred from a linguistic point of view, but also because it is cheaper compared to the other options, the lexical coverage is much wider, and it can be customized and updated.

4.3 Evaluating Option 3's raw output

We then started the third phase of the testing process. This test was aimed at analysing and scoring the accuracy and fluency of the raw output and validate the results of the automatic scoring (Marg, 2016). It also allowed us to identify the typical issues in the MT output, and to start putting together post-editing instructions. The test was performed by two linguists for each language pair (2 for Italian into English and 2 for English into Italian).

4.3.1 Results for English into Italian

The test for English into Italian was performed by two native in-house translators specialized in financial translation. The text translated was a piece of Transfer Pricing document. Both linguists scored accuracy and fluency consistently.

As for accuracy, the major issues concern mistranslations (calques, antonyms, positive to negative sentence or vice versa), omissions (especially missing numbers) and terminology.

As for fluency, there seemed to be a shared opinion as to the grammar mistakes (gender and number agreement, wrong and/or missing prepositions, *consecutio temporum*, translation of modal verbs) and locale adaptation (numbers and measurements were not localized).

4.3.2 Results for Italian into English

The test for Italian into English was performed by two external preferred native translators who specialize in financial translation (Plitt and Masselot, 2010). The text translated was a piece of Financial Statements. Also in this case, both linguists scored accuracy and fluency consistently.

As for accuracy, major issues concerned mistranslations (proper nouns and acronyms replaced by random words) and omissions. Terminology also appeared to be problematic, while numbers were not missing in MT output for this language pair.

As for fluency, major issues concerned word order (which often mirrors Italian word order),

grammar (primarily verb tenses) and locale adaptation (numbers and measurements were not localized).

To sum up, some types of issues were spotted in both language pairs, while others were language pair specific.

5 MT into Production: Preliminary Phase

Having chosen the preferred MT engine and identified the main potential issues, we decided to run a few more tests to analyse more deeply the mistake trends for each language pair. The results of the analysis showed that, to achieve publishable quality – required for our business – a full-post editing was necessary. Originally, to reach the required quality, we implemented a TP process which envisaged a first step (Translation) and a second step (Proofreading). In the MT pilot projects, we decided to keep two steps to ensure top quality: post-editing and review – basically the translation step was replaced by post-editing.

5.1 Onboarding Freelance Post-Editors

We then started thinking of the new workflow and how it would merge with our existing vendor base. As mentioned above, it did not include any experienced post-editors. For this reason, we started organizing non-mandatory Machine Translation Post-Editing training for the suitable vendors in our database (Massardo et al., 2016). We sent them an invite and explained them that we were implementing MT in our workflow and they would be offered MT post-editing tasks in the near future.

We gave three training sessions: one in 2018 and two in 2019. The training we gave in 2018 was a generic MTPE training and applied to all language pairs and domains, while of the two training sessions we gave in 2019 one was focused on post-editing our engine's raw Transfer Pricing translation from English into Italian, and the other was focused on post-editing our engine's raw Financial Statements translation from Italian into English in a CAT tool environment. Apart from covering the topics already discussed in the 2018 edition, the 2019 training also focused on the most frequent mistakes delivered by our engine and on how MT was introduced in the workflow. In

March 2019 our database included a number of new post editors specialized in finance.

5.2 Instructions for Linguists

We decided to create an instructions file for linguists to be sent over with each project, in order to remind them the guidelines for full post-editing³ and the most frequent known engine errors (Joscelyne, 2008).

Apart from indicating the above-mentioned most frequent mistakes for each language pair, instructions warned linguists about some strange errors delivered by the engine in very short strings only, which are always a challenge for MT. Basically, proper nouns (company names, cities...) and acronyms (GBP, HPC...) are frequently replaced by random words. Sometimes errors delivered by neural MT engines cannot be fixed by implementing changes in the engine directly, so for the time being we decided to mention this issue in the instructions.

The instructions file also included other useful key take-aways, suggestions and reminders on how to perform post-editing and review of postedited content in a CAT tool environment, like how to understand if the translation of the segment comes from TM or MT, and the indication to follow the TM as for preferred date and currency format.

6 MT into Production: Pilot Phase

As mentioned above, projects with PE are handled the same way of the standard TP projects, with translation being replaced by post-editing, plus three new steps: pre-editing, pretranslation with MT and post edit distance measurement, all performed by our internal staff.

After having ascertained that a project is suitable for MTPE by following some internal criteria, we start the pre-editing step, which consists of some minor interventions on the source file to facilitate machine processing, like running a spell-check and removing double spaces. We then upload the file on the CAT tool and pretranslate number-only and untranslatable-only segments. Number-only segments are automatically localized by the CAT tool – this way we reduce the risk of having them mistranslated or

³ As explained in the ISO 18587:2017 standard, which provides guidelines for the process of full, human post-editing of machine translation output.

not localized. Untranslatable-only segments are short segments made up by proper names, company names, acronyms (like EBIT, EBITDA) or characters (% , € , - ...) that do not change from Italian into English and vice versa.

Then, the file is pretranslated with the project TM: the TM is leveraged down to 75% Fuzzy matches. Anything that does not have a $\geq 75\%$ match in the TM is considered a New Word segment, that is sent to MT. This threshold was chosen because our internal linguists did not find Fuzzy matches below 75% to be very useful as a starting point for translation. For the time being, we like forcing post-editors to work on machine translated content, so, to reduce the temptation of writing the translation from scratch, we pretranslate with MT the matches below 75% and New Word segments. The file is then sent to post-editors along with instructions, and after post-editing is complete, the reviewer, who also received the instructions, can start working on the file.

Since one of the most delicate issues regards key terminology consistency, we associate an empty glossary to each project. Linguists are required to populate it during the translation step with key terminology which is translated wrongly or inconsistently by the MT engine. CAT tool's QA check – which is set up internally upon project creation – will automatically deliver an error message every time a term in the glossary is not translated properly. These glossaries will also be used to update the MT engine.

After the translation is delivered to the customer, we run an auto scoring test on the MT post-edited segments only to see how much of the raw output was changed, and send a survey to the linguists so that they can express their opinion on the quality of the raw output and provide suggestions for improvement. These last two steps are extremely useful for the future updating of the engine.

So far, we have run a fairly big amount of pilot projects, and the results in terms of productivity increase for both language pairs are satisfactory. Productivity increase varies depending on many factors – vendor's experience with MTPE (Guerberof, 2009) and individual speed, source file, language pair, client requirements, etc. – but on average it ranges between 20-25%. We expect this percentage to increase after engine updating and fine-tuning. We ensured that the quality of the

final translation was of the same high standard as before by running the same QA processes.

7 Vendors' Feedback

Translators often get stressed when they hear the word “machine translation”, especially the ones who are more reluctant to try out new technologies. As mentioned above, many translators of our vendor base are not familiar with CAT tools and are not willing to learn how to use them. Some of them are so experienced, productive and used to work “the old way” that they see anything technologically new as something that will affect them negatively.

To sum up, our background was not really the most suitable within which to implement MT, however, some of the freelancers were happy and curious to take part to the training and the pilot projects. They wanted to start getting familiar with machine translation, since more and more LSPs are implementing it in their workflow – this therefore means more job opportunities for them, as well as an increase in productivity.

In all the training sessions we gave, resources asked how the implementation of MT will affect their rates. We were expecting a lot of concerns on this matter (this was also flagged by O'Brien et al., 2009), so we decided to keep the rates unchanged during the pilot phase – basically matches pretranslated with MT were paid like new words for all projects. This way we managed to convince many of them to give MT a try.

As mentioned earlier, a short questionnaire was sent to all the vendors who took part to the pilot projects in order to gather feedback and suggestions. The answers show that the greater part of them feels the MT output was overall useful as a starting point for their translation, and that in most cases they used big portions of MT raw output, introducing minor changes only. Many of them stated they felt they worked faster thanks to MT and that they are willing to work on more MTPE projects. Besides, they were left some blank space where to add suggestions and a description of the most common issues they found in the raw output. Among the issues flagged by linguists, apart from the ones already discussed, the problem of aligning with client's preferred format for amounts and dates was raised, as well as the lack of creativity of the engine and misinterpretation of the meaning of some sentences.

References

- Castilho, Sheila, Joss Moorkens, Federico Gaspari, Iacer Calixto, John Tinsley and Andy Way. 2017. Is Neural Machine Translation the New State of the Art? *Prague Bulletin of Mathematical Linguistics*, 108.
- Guerberof, Ana. 2009. Productivity and quality in MT post-editing. *Proceedings of MT Summit XII - Workshop: Beyond Translation Memories: New Tools for Translators MT*, Ottawa, Canada.
- Joscelyne, Andrew. 2008. Post-editing: Update on Best Practices, *TAUS Report*.
- Marg, Lena. 2016. *The Trials and Tribulations of Predicting Machine Translation Post-Editing Productivity*. Presented at the 2016 Language Resources Evaluation Conference (LREC). http://www.lrecconf.org/proceedings/lrec2016/pdf/810_Paper.pdf Accessed 16 April 2019
- Massardo, Isabella, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri and Katrin Drescher. 2016. *MT Post-Editing Guidelines*. TAUS Signature Editions.
- O'Brien Sharon, Johann Roturier, Giselle de Almeida. 2009. Postediting Machine Translation Output Guidelines. *Proceedings of MT Summit XII – Tutorials*, Ottawa, Canada.
- Plitt, Mirko and François Masselot. 2010. A productivity test of statistical machine translation post-editing in a typical localisation context. *Prague Bulletin of Mathematical Linguistics*, 93:7– 16.
- Schmidt, Tanja and Lena Marg. 2018. How to Move to Neural Machine Translation for Enterprise-Scale Programs—An Early Adoption Case Study. *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, Alicante, Spain.
- TAUS - Enabling better translation. 2012. Advancing Best Practices in Machine Translation Evaluation, *TAUS Labs Report*.