ROI Analysis model for Language Service Providers

Ekaterina Stambolieva

euroscript Luxembourg S.à. r.l. 55 rue de Luxembourg L-8077 Luxembourg

ekaterina.stambolieva@euroscript.lu

Abstract

Return on Investment (ROI) analysis plays a primary role in business strategic planning; however it is not such a straightforward process in itself. From the perspective of a Language Service Provider's (LSP) case study, an economic impact analysis of machine translation (MT) technology usage is required. Potential initial development and customization costs are investigated, as well as estimations of financial consequences resulting from the use of MT in professional translation. Different courses of action are weighed up, and supporting arguments are presented for the strategy of choice.

1 Introduction

There is a current need for accurate ROI analysis schemes for LSPs. As Wiggins and Vashee (2012) discuss, LSPs that have not incorporated MT in their internal workflow by 2015 will be facing the risk of bankruptcy. Moreover, both researchers and professionals agree it is out of the question whether LSPs should implement the usage of MT in their respective businesses or not. The question has shifted to how LSPs can profit from the usage of MT internally and harvest the proven benefits.

ROI calculation models for LSPs are not only required to include predicted sales rates of MT translation, but also the customer relationship process. This combines additional tasks such as proofreading, document conversion, and revision, among others. Therefore, revenue is not the only factor relating to translation impacting the company's financial gain. The aim of the paper is to deliver a financial comparison between different internal MT solutions within a LSP. It also aims to propose a winning business strategy based on the difference in magnitudes of gain and cost for three possible courses of action. The first proposed course of action is against the use of MT, when human translation of incoming documents is provided to clients. The second option discusses the delivery of post-edited ma-chine translation output to customers using an open-source MT decoder. The final option leans towards the delivery of post-edited MT translation using a commercially-distributed MT tool, implying that no internal development is involved. The argument proposed is that the ratio between the magnitude of investment gain to the magnitude of investment costs (ROI calculation) increases when post-editing output of a customized opensource MT decoder is delivered to clients. An open-source solution is preferred, based on low initial investment cost. An estimation of the MT output is possible when the development of MT models is concentrated in-house. Consequently, it is suggested that when investing in MT development in-house, the output is more predictable and reflects better on the LSP's standards of translation quality than when purchasing a licensed MT solution.

LSPs, as a specific type of an organization, have an inconsistent flow of clients but continuous demand for translation services. Due to this, expected revenue prediction is challenging. When engaged internally with MT development, a LSP cuts costs on obtaining a license for other available solutions. Excluded costs are fees for third party provision of translations. Included costs are development costs and hardware purchase. It is expected that the LSP does not share internal development costs with its clients; therefore there is a need for a more accurate ROI analysis model tuned to the LSP's specific needs and requirements. With open-source MT solutions internal costs of implementing MT systems decrease. The initial investment cost is lower than when buying MT licensed solutions. No benefits are shared with another MT solution provider. No external dependencies are present and the business controls entirely the MT development course of action.

Reflecting on all of the above, a one can suggest a model for ROI calculations for a LSP, which is described in Section 7 Comparative analysis. Section 2 defines ROI in a LSP context. Section 3 discusses related literature and relevant findings from a competitor's perspective. Section 4 concentrates on costs and benefits arising from these outlays. Section 6 gives an interesting insight into unpredictable challenges such as vendor pricing model and its impact on the business. Finally, Section 7 concludes the paper.

2 Return On Investment

As discussed briefly in the Introduction, ROI is the ratio between the investment gain and cost magnitude. According to DePalma and Hills (2010) ROI "is a calculation of the benefits that result from spending money or allocating resources. ... Other organizations measure their return in terms of pumping up brand awareness, gaining market share, saving money, or reducing support calls". The ROI formula follows as shown in Figure 1.:

$$ROI = \frac{(Investment \ Gain - Investment \ Cost)}{Investment \ Cost}$$

Figure 1: LSP ROI formula

In an LSP-context ROI is measured in relation to stakeholders' satisfaction, reduced time of translation, customers' added value such as accelerated time to market (TTM) in international markets, among others. Most importantly, a LSP would calculate investment gain in terms of reduced costs of the internal translation and document delivery process. When MT contributes to faster translation delivery for a lower price, the LSP, following the same document processing tasks and delivering the same order, is capable of producing more for the time required before MT incorporation in the daily business.

3 Relevant Literature

Not much relevant literature can be found on the specific topic of estimating ROI in a LSP context. Nonetheless, different language and translation technology professionals have invested time and effort into discussing business and pricing models, which shed more light on ROI calculation. Sojnóczky (2013) emphasizes on margin calculation and stakeholders' satisfaction in the case when MT output is post-edited and delivered to the clients. Sojnóczky, the managing director of Hunnect¹, implements a business model, which concentrates on margin shrinking benefits and its results are directly projected to clients' and translators' satisfaction. Sojnóczky (2013) does not discuss LSP costs for internal MT development. He emphasizes on margin shrinking, which is directly linked to investment gain. Hunnect's translation provision process involves three tasks - translation, bilingual editing and proofing. The time estimation of these tasks per order, shown in percentage, is the following: translation requires 50% of the time, editing -20%, proofing -5%, and the rest 25\% is the marginal buffer time. The time management planning for the post-edited MT output is as follows: 20% of the time is devoted to translation, 30% - for post-editing (PE), 5% for proofing. The other 45% of the time is the margin time until delivery. In reality, Sojnóczky (2013) observes the margin time left after translation, postediting and proofing for Hunnect's case is 50%. Therefore, the use of MT in the translation process reduced the delivery time or 35% savings of time. 35% savings of time results in ... quicker potential ROI. The vendor pricing scheme implemented by Sojnóczky (2013) is that postedited MT is paid 60% of the original price. Consequently, the company cost for delivering an order is reduced by 40% and time to delivery is increased by 35%. As mentioned in Section 7, managers need to be careful when developing vendor pricing model strategies as it is a sensitive matter. Additionally, Sojnóczky (2013) observes 68% increase of productivity among trained to post-edit translators, which reflects directly the translators' pay. The statistics Sojnóczky (2013) observes show that the translators' pay has increased with at least 1% instead of decreasing as feared.

¹ http://www.hunnect.hu/

Opposed to Sojnóczky (2013), McMahon (2009) estimates the post-editing costs of a commercial business for instant communication messaging translation. Interestingly, his cost estimations cover corpus management and acquisition, 5 to 200 GBP per year, customization costs, human resources training, 5 to 10 GBP per year, among others. He measures investment cost of product licenses to 100 GBP per year. Overall, the total investment costs are estimated to 180 GBP per year. The predicted ROI for the first year after successful MT implementation equals to minus 30000 GBP, therefore is negative speaking in financial terms. The vendor pricing models is estimated to 65-85% of the original pay per word. What is intriguing is the fact benefits are not only measured in money. McMahon (2009), as Sojnóczky (2013), measures the user satisfaction rate, which in the case of Lionbridge² has increased with 30-50%. The success rate measured varies between 5% and 25%.

Wiggins (2013) gives yet another perspective of investment and MT implementation in Asia Online³ business environment. Some of the advantages of providing MT-incorporated solutions Wiggins (2013) list are reduced translation costs, faster delivery time, expansion of existing relationships with clients, broadening offered functionality and opening new market possibilities.

Hernández-Lasa (2011) reports intriguing findings in a Sajan's⁴ case study presentation. In Sajan, achievements due to the incorporation of MT in the translation process reach to 60% cost reduction and savings of 77% time to translate compared to the results of their customer's proprietary MT solution results.

4 Costs and benefits

In this section, the focus is drawn on LSPcentered estimation of costs for the three scenarios suggested in the Introduction. The three scenarios are the following: no incorporation of MT in the daily business of a LSP, incorporation of internally developed MT solution, and incorporation of a licensed solution for MT in the translation process. All possible costs for each scenario are drawn and listed in tables. At the end of the Section, a cost comparison is attempted.

Table 1. gives an overview of expected costs in translation. It is important to emphasize that the total cost of number of translated segments varies between vendors and language direction as the pricing models are adapted to fit best the expected translation effort.

Table 2. expands the overview of costs to cover the whole translation process depending on the different services a LSP offers to its clients. Scenario 2 and 3 share some of the listed costs in Table 2, but include additional ones. The total cost of resources, for instance, includes human resources costs, hardware and software license costs, among others.

Cost of Human Translation Abbreviation

~	~
Cost of Number of Translated Seg-	Co _{segm}
ments	-
Cost of TM Segment Matches	Co _{tmsegm}
Cost of Bilingual Editing	Co _{edit}
Cost of Proof-reading	Coproof
Cost of Human Translation	Co _{tran}

Table 1: Cost of Human Translation

Abbreviation

Cost of Translation Process

Cost of Document Conversion	Co _{conv}
Cost of Project Coordination	Co _{proj}
Cost of Specific Services ⁵	Co _{serv}
Cost of Translation	Co _{tran}
Cost of Terminology Management	Co _{term}
Cost of Resources	Cores
Cost of Translation Process	Co _{trproc}

Table 2: Cost Translation Process

As LSPs rely on technology, the minimal requirements in place are the presence of a CAT tool system, and a project coordination framework. Personnel training on CAT tool usage as well as project coordination usage induction are not regarded as costs in the current scenarios, as they are expected to have been completed. However, post-editing training and MT development

² http://www.lionbridge.com/a

³ http://www.asiaonline.net/

⁴ http://www.sajan.com/

⁵ such as Apostile, Courier Services, Layout services, Certification, Urgency fees

training, among others, are considered as investment costs since they depend on the new involvement of MT technology in the business processes. Therefore, the total LSP investment cost for providing pure human translation services is presented in Figure 2:

$$InvCOST_1 = Co_{trproc} = Co_{conv} + Co_{proj}$$

$$+Co_{serv} + Co_{tran} + Co_{term} + Co_{res}$$

Figure 2: Scenario 1 Investment Cost Formula

Table 3. aims to provide cost overview of translation within the context of Scenario 2, namely when a LSP incorporates in-house developed MT solution's output in the translation delivery process. The emphasis comes down to the cost of the number of translated segments, which is marked with Comtsegm instead of Cosegm as in Table 1. For good language and translation models, when overall MT output BLEU score (Koehn 2010) is higher than 65%, a MT pre-translated segment translation is paid 50% of the human translation rate per segment. Consequently, $Co_{mtsegm} = 50\% * Co_{segm}$. To our knowledge, our internal proofing costs do not change according to recent statistics. With time, these statistics can change and guide better pricing model development.

Cost of Human Translation	Abbreviation
Cost of Number of Translated Seg-	Comtsegm
ments (Cost of Post Editing)	
Cost of TM Segment Matches	Co _{tmsegm}
Cost of Proof-reading	Co _{proof}
Cost of Human Translation	Co _{mttran}

Table 3: Cost of Translation when MT solution is implemented internally

Table 4. shows the translation process investment costs when MT is developed internally on LSP premises. The document conversion, project coordination, additional services costs, along with the cost of terminology management, and resources, are identical and the same as in Scenario 1. It is interesting to note the costs that accompany internal MT development. These costs include but are not restricted to customization,

automation, software and hardware costs. A cost for trainings is introduced. It covers project coordinators MT training expenses. Additionally it includes MT specialist new methodology and approach trainings. These costs can be roughly estimated to the expenses incurred for the number of employees required per day. Depending of the scale of the MT projects, hardware costs for two servers allowing simultaneous translation requests can easily reach 60K euros. The hardware cost is repetitive every year as more projects and users are expected to benefit from MT. The cost of customization and optimization include human development effort and is measured in man-hours. This cost is directly linked to the software licensing expenses as licensed tools are used to accelerate the MT specialists' development effort. More managerial tasks by human resources are required such as risk management monitoring and development of prevention plan strategies.

Nevertheless, an important observation to be mentioned is that the additional investment costs do not need to be made when kicking-off an

Cost of Translation Process	Abbreviation
Cost of Document Conversion	Co _{conv}
Cost of Project Coordination	Co _{proj}
Cost of Specific Services ⁷	Co _{serv}
Cost of Translation	Co _{mttran}
Cost of Terminology Management	Coterm
Total Cost of Resources	Cores
Additional Costs for Scenario 2	
Cost of Trainings	Co _{train}
Cost of Customization (time to	Co _{cust}
quality)	
Cost of Development	Co _{dev}
Cost of Automation	Co _{aut}
Cost of Hardware	Co _{hard}
Cost of Third Party Software Li-	Colic
cense	
Cost of Risk Management	Co _{risk}
Cost MT Specialists Human Re-	Co _{spec}
sources	
Cost of Translation Process	Codevmt

Table 4: Cost of Translation Process when MT solution is implemented internally

⁷ such as Apostile services, Courier services, Layout services, Certification, Urgency fees

⁶ equal to 50% *Cosegm

internal for the LSP MT development project. This scenario suggests high investment costs, which are distributed over time, and potentially leads towards high long-term ROI perspective. As McMahon (2009) shows first year ROI ratio is negative. When implementing this scenario, no fees for third party provision of MT output need to be foreseen. No MT solution license cost is added, in return development, customization and automation expenses need to be estimated and taken into consideration.

Cost of Translation Process

Abbreviation

Cost of Document Conversion	Co _{conv}
Cost of Project Coordination	Co _{proj}
Cost of Specific Services	Co _{serv}
Cost of Translation	Co _{mttran}
Cost of Terminology Management	Co _{term}
Cost of Resources	Co _{res}
Additional Costs for Scenario 2	
Cost of Trainings	Co _{train}
Cost of Customization (time to	Co _{cust}
quality)	
Cost of Development	Co _{dev}
Cost of Automation	Co _{aut}
Cost of Hardware	Co _{hard}
Cost of Third Party Software Li-	Colic
cense	
Cost of Risk Management	Co _{risk}
Cost MT Specialists Human Re-	Co _{spec}
sources	
Additional Costs for Scenario 3	
Cost of Licensed MT Solution	Co _{extmt}
Cost of MT service fees ⁸	Co _{fee}
Cost of MT client fees	Co _{effe}
Cost of MT provision fees	Coprov
Cost of Translation Process	Colicmt

Table 5: Cost of Translation Process when MT solution is licensed

The investment cost estimation formula is for scenario 2 is shown in Figure 3, $InvCOST_2 = Co_{devmt}$. It accumulates all costs mentioned in Table 4. It is higher that $InvCost_1$, nevertheless in scenario 2 MT benefits and revenue are expected in the long-term perspective. Scenario 1 costs are

considered as the minimum costs a LSP has in order to provide services to its clients.

$$InvCOST_{2} = Co_{mttrproc} = Co_{conv} + Co_{proj}$$
$$+Co_{serv} + Co_{mttran} + Co_{term} + Co_{res}$$
$$+Co_{train} + Co_{cust} + Co_{dev} + Co_{aut}$$
$$+Co_{hard} + Co_{lic} + Co_{risk} + Co_{spec}$$

Figure 3: Scenario 2 Investment Cost Formula

The costs of human translation for the scenario 3 are the same as in scenario 2, which are listed in Table 5. However, additional costs must be taken into account when licensing a MT solution from a third party provider. These costs include license cost and fees for distribution of the MT output. On the positive side, no development and hardware costs are relevant for this scenario:

Therefore, the formula shown in Figure 4., , $InvCOST_3 = Co_{licmt}$, is used as a basis when estimating LSP's investment cost when incorporating licensed MT solution output in translation:

$$InvCOST_{3} = Co_{mttrproc} = Co_{conv} + Co_{proj}$$
$$+Co_{serv} + Co_{mttran} + Co_{term} + Co_{res}$$
$$+Co_{train} + Co_{cust} + \frac{Co_{dev}}{Co_{dev}} + Co_{aut}$$
$$+\frac{Co_{hard}}{Co_{hard}} + Co_{lic} + Co_{risk} + Co_{spec}$$
$$+Co_{extmt} + Co_{fee} + Co_{efee} + Co_{prov}$$

Figure 4: Scenario 3 Investment Cost Formula

The investment costs of scenario 2 and 3 are much higher than the investment cost of scenario 1. As mentioned before, no MT benefits can be expected in scenario 1 due to MT implementation in processes absence. Scenario 2 costs per year can be greater to scenario 3 costs. Nevertheless scenario 2 costs are distributed throughout the year and little investment cost is required to kick-off MT related projects. Scenario 3 requires license fee payment done before MT kick-off start, implying no MT benefits can be seen before paying for MT. In scenario 2, MT benefits

⁸ Where the service fees are applied per engine, they are multiplied per 3 when 3 translation engines are in place

can be assessed with small initial investment cost in place.

Figure 5. gives a suggestive calculation model for LSP's investment gain when following scenario 2. The direct benefits of MT incorporation in the daily translation business for the LSP are lower translation cost and shrunken margins, InvGAINtr. The potential benefits of the use of MT in the LSP's translation process (Scenario 2) include faster translation provision. The faster translation services build better client relationships and increase client satisfaction. Another potential benefit directly linked to the first one is that a client may commit to work with the LSP on more language directions and pairs. Therefore, the LSP broadens its range of offered language direction translations, and covers better existing markets, InvGAIN₂. When working with freelance translators, these benefits can be harvested in the case a strong freelancers' commitment to MT is achieved. If the freelancers lose money because of MT, a risk the LSP faces is their rejection of MT related tasks. If MTincluding orders are not accepted to be worked on, the LSP's possible financial gain from MT transfers to LSP's possible financial loss. Moreover, the LSP cannot license its MT solution to other LSPs or companies. Therefore it cannot position itself on the MT provision market place.

$$InvGAIN_2 = G_{tr} + G_{newlg}$$

Figure 5: Scenario 2 Investment Gain Formula

The potential benefits of the internal development of MT within a LSP, include all but are not limited to the potential benefits from scenario 2. A potential benefit is LSP positioning in a MT provision market niche for one or more language directions of translation. As new markets bring new revenue, the LSP's investment gain is expected to rise when new projects are acquired. Figure 6. presents a suggestive LSP's investment gain calculation, InvGAIN3, based on the aforementioned:

$$InvGAIN_3 = G_{tr} + G_{newMark} + G_{newlg}$$

Figure 6: Scenario 3 Investment Gain Formula

5 Comparative Analysis

The internal MT development at euroscript is still young. Our preferred course of action is described in scenario 2. Nevertheless, yet more statistics are to be collected and analyzed before we publish actual investment cost and gain figures. Thus we do not report the number behind our ROI calculation. Instead, we opt for empowering other LSPs by providing a ROI calculation model, which is LSP-centered. A ROI calculation model for scenario 1 is irrelevant for the current analysis as it does not include MT incorporation in the translation business.

The ROI calculation model suggested for scenario 2 is the following:

$$ROI_2 = \frac{(InvGAIN_2 - Co_{devmt})}{Co_{devmt}}$$

Figure 7: Scenario 2 ROI calculation

The ROI calculation model suggested for scenario 3 is the following:

$$ROI_{3} = \frac{(InvGAIN_{3} - Co_{licmt})}{Co_{licmt}}$$

Figure 8: Scenario 3 ROI calculation

As a reference, Section 4 gives detailed explanation in details on how the investment gain and cost for each of the two scenarios is calculated.

What this analysis does not include is an assessment of volumes required to make a financial comparison between human translation and postedited MT gain.

6 Vendor Pricing Model Challenges

A difficult challenge to face is designing a winning vendor pricing model. Even in the cases of google⁹-like machine translation output, it is arduous to benefit from vendor involvement with MT. The reason why originates from the fact that when incorporating MT output in the translation process workflow, per segment rates are directly affected and reduced (the reductions can reach 50% of the normal price). From vendors' point of view, MT threatens to reduce per segment rates,

⁹ http://www.google.com/

which in return leads to hourly rate destabilization. Regardless of MT output quality, vendors tend to decline jobs due to the fact MT output is incorporated in pre-translated documents. This decrease of vendor involvement leads us to a conclusion that purely showing good MT results is not sufficient and more work is required to be done in order to start gaining investment back and increasing the ROI ratio. What we discovered on our own, stated additionally by Sojnóczky (2013), is that human translators' engagement is MT development is vital. People are the key factor of success of every business and winning strategies must actively involve employees on different organizational levels in technological (including MT) solution processes. In euroscript¹⁰, we regularly ask for in-house translators' MT evaluation feedback in terms of different categories of MT output errors. We devote time to correcting these mistakes, informing the in-house translators of the improvements made and engaging ourselves into increasing translators' satisfaction with working with MT. We consider different options of post-editing training focused on translators following the strategy of involving translators more with MT. This idea is also supported by: Hernández-Lasa (2011) and Wiggins (2013). Our preliminary results show that roughly 50% of our in-house translators, who work with MT, consider its understandability as good opposed to acceptable or bad (see Avramidis et al., 2012; Vilar et al., 2006 for MT features that influence MT understandability). The other 50% consider it as bad. Moreover, the in-house translators classify the post-editing effort (Avramidis et al., 2012) as intermediate, but this information might be unreliable due to insufficient numbers of answers on the MT output understandability question.

The feedback we collect from vendors shows a different picture. Some vendors refuse to work on documents, which incorporate MT. We regard their unwillingness to be involved with MT as evaluating MT output understandability as bad. A feedback comment we received is that MT output requires more time to be corrected compared to translation from scratch. Such negative feedback leads to concluding post-editing training is needed in order for the translators to work productively with machine translation. Our pre-liminary results show 11% freelance satisfaction

with the understandability of the MT output. Another 33% rate the understandability as acceptable, and the rest, 56%, regard it as bad or incomprehensible.

We realized that in the context of our business, we struggle to communicate what the MT terms quality, understandability and post-editing effort stand for. In our results it is clearly visible that professionals in different fields of application consider these terms differently. For example, a translator and a linguistic solutions architect regard the same snippet of translated text as good and bad. The translator is concerned with the translation quality of the MT output, which is bad for the reason that 20-25% of each segment must be changed. For the linguistic solutions architect these results show good performance of the MT engine and satisfactory results.

An idea worth considering is that the vendor pricing models need to be language pair and language direction sensitive. As each MT engine's output confidence score estimate is different for different languages, pay per segment/word need to be sensitive to this in return.

The identified challenge lays within the fact a clear communication scheme of expected results and performance should be developed and deployed. Post-editing training is to be offered to all involved parties, including in-house and free-lance translators.

7 Conclusion

This paper discusses an attempt to delineate a LSP-centered ROI analysis strategy. Estimation of costs for in-house development of machine translation engines is suggested along with estimation of costs for licensing an off-the-shelf MT solution, which is incorporated in an LSP. Arguments are given to support the use of technologies in translation, in particular – machine translation. An initial attempt to suggest a LSP-centered ROI calculation is presented. A more elaborate descriptive analysis of investment gain is progress. Its public release is expected to ease LSP professionals in the ambiguous task of estimating MT investment ROI.

References

Avramidis, Eleftherios, Aljoscha Burchardt, Christian Federmann, Maja Popovic, Cindy Tscherwinka and David Vilar Torres. 2012. Involving Language Professionals in the Evaluation of Machine Translation, In 8th ELRA Conference on Language Re-

¹⁰ http://www.euroscript.com/

sources and Evaluation, European Language Resources Association (ELRA), pages 1127-1130, Istanbul, Turkey.

- Binger, Gregory, and Dion Wiggins. 2012. The Evolution of Translation: What LSPs need to know to survive and prosper in the rapidly evolving automated translation world. Paper presented online, 22nd May 2013.
- DePalma, Donald, and Mimi Hills. 2010. Localization Return on Investment: Quantifying the Value of Localization in High-Tech. Common Sense Advisory, Inc. Published April 2010.
- Koehn, Phillip. 2010. Automatic Evaluation. *Statistical Machine Translation*, Cambridge University Press, pages 226-228.
- Hernández-Lasa, Iñaki. 2011. Sajan's Case Study. Proceedings of the 16th Annual LRC Internationalization and Localization Conference & 3rd Multilingual Web Workshop, 21-23 September 2011, Limerick, Ireland.
- McMahon, Nicholas. 2009. Lionbridge Technology Overview: Technology Best Practice. Paper presented online, 15th July 2009.
- Sojnóczky, Saándor. 2013. Hunnect's Use Case. *TAUS Machine Translation Showcase at Localization World*, Singapore, 10th April 2013.
- Vilar, David, Jia Xu, Luis Fernando D'Haro, and Hermann Ney. 2006. Error Analysis of Machine Translation Output. *International Conference on Language Resources and Evaluation*, pages 697– 702, Genoa, Italy.
- Wiggins, Dion, and Kirti Vashee. 2012. MT & ROI Scoping, Defining, Measuring and Selecting Suitalbe Projects. Paper presented online, 29th November 2012.
- Wiggins, Dion. 2013. Business Strategies for Building Strategic Advantage and Revenue from Machine Translation. *Taus Machine Translation Showcase at Localization World*, Singapore, 10th April 2013.