Machine Translation for enterprise technical communications – a journey of discovery

Morgan O'Brien mobrien@mcafee.com
Ana Duarte ana_duarte@mcAfee.com
Global Language Services, Intel Security, Cork, Ireland

Abstract

There has always been a high quality requirement from large corporations regarding machine translation due to perceptions and common beliefs which in turn makes it difficult to break into the area without well-maintained engines and processes. This project details from the combination of various internal efforts towards automation in translation of technical communications. Namely working with external language providers/partners, testing the offerings, understanding the nature of our source content and the inclusion of machine translation technologies for rolling out Post-Editing Machine Translation (PEMT).

1. Credits

This paper is derived from research on tools, technology and processes since 2012 towards the implementation of Machine Translation (MT) within the Localization workflow of Intel Security. It takes input from machine translation service providers, translation vendors and post-editors, language quality team, localization professionals within the company and the various departments where machine translation helps their productivity and allows them to reach their target customers and a wider audience.

2. Introduction

If one was to listen to sales pitches from various MT providers, one could learn a lot. It is noticeable that many claim their system is better in some unique way when compared to other systems or the general knowledge in the area. Too much focus on the individual selling points will distract the receiver from possibly more important pitfalls of the rollout process. To separate the jargon from the relevant, one needs to take a step back and look at content from its conception to its consumption and analyze the supply and demand of it.

Quite often, the place to start isn't on the MT, but the internal content, the tools and the people: the three main pillars we need to "shape" in order to get acceptance for MT and effect whatever systems are needed. We selected 5 different content types; Product Documentation, Knowledge Base, Community generated content, Global Definitions database and Product UI. Our main focus was on Knowledge Base articles and Product Documentation as these were two areas where large corpus existed for MT training and structured authoring teams were in place. Below we will follow the discovery process on our initial testing of content to MT. We will detail some of the tests and the systems that need to be altered and provides some recommendations based on the lessons learned. The main output from this is to be our MT strategy to a PEMT rollout.

3. Source, Target and Speed

The first time people start learning about MT there can be a lot to digest. Confidence scores are quoted as unique selling points, BLEU scores are proudly displayed as a metric of quality. MT providers offer pre-ordering, automatic post editing and domain adaptation which they say increases their quality, and there can be decisions to be made around Statistical Machine Translation (SMT), Rules Based Machine Translation (RBMT) or Hybrid methods which are a combination of systems. It can't be denied that all these things play a part as there are studies that prove this. But how much impact do they really have effecting the ability to roll out high quality MT? Where does attention need to be paid, and what are the priorities? It can be hard to tell at the beginning of an MT program.

Having gone through the required ramp up of technical knowledge with the many great online resources and taken advantage of talking to peers in the industry, it comes to a point when actions need to be taken and one must choose a method to proceed with. The focus for our tests was on PEMT and 3 areas stood out for measurement; Productivity, Target Language Quality and Source Language Quality. These were largely driven by internal requirements for Cost, Speed and Quality.

It was important we looked at Productivity as demonstrated in Post-Editing studies like Plitt, M., & Masselot, F. (2010) where post-editing substantially increased their productivity when compared to Human Translation (HT). This paper the authors talk about productivity measured in Time, and this seemed a reasonable starting point. We then found a number of tools available to do this such as iOmegaT where timing data is measured in a desktop translation tool. One of the advantages of this tool was the ability to track time per segment, but also revisits to segments, which give a clear picture of how much effort was given to each Post-Edited segment. It would be considered normal practice in Translation and Post-Editing for the translator to revisit segments once context became clearer while translating similar segments. Also as iOmegaT is a desktop Computer Aided Translation (CAT) tool, it is closer to the native working environment for translators reducing that variable from the tests. We also mixed the post-editing task by creating a project TM with segments to be fully translated (no MT) and also inserted some previous TM matches for segments to be leveraged/reviewed. The final measurement we wanted from this was how many words were Post-Edited in 1 day relative to how many were translated using the same environment and project. Again, "words per day" is a standard metric in our business for forecasting translation time in projects so it was important to leverage something that is already generally understood. We displayed the throughput relative to the 4 final engines we were testing (4 languages). The final throughputs (Fig. 1) were recorded where the minimum quality bar was met (80% pass mark for LQA).

	#1	#2	#3	#4
Doc	4982	7292	8826	5455
KB	4743	18262	4461	5176

Figure 1: Throughput per day (words) for Doc and Knowledge Base content

We needed to measure quality of target language and for this we already had Language Quality Analysis steps in place. Our LQA score is another measurement that exists in our business day to day, so again it's something that people in the company already understand. This quality measurement is based on the LISA LQA model and the results are in the form of a chart with an overall score out of 100. While the LISA LQA model will do for general quality assessment we also needed something more specific and repeatable. For this we complimented LQA with an Edit Distance measurement on every segment, giving us some drilldown data when investigating problematic segments later on. We displayed this on a graph to demonstrate where most of the effort was for Post-Editing.

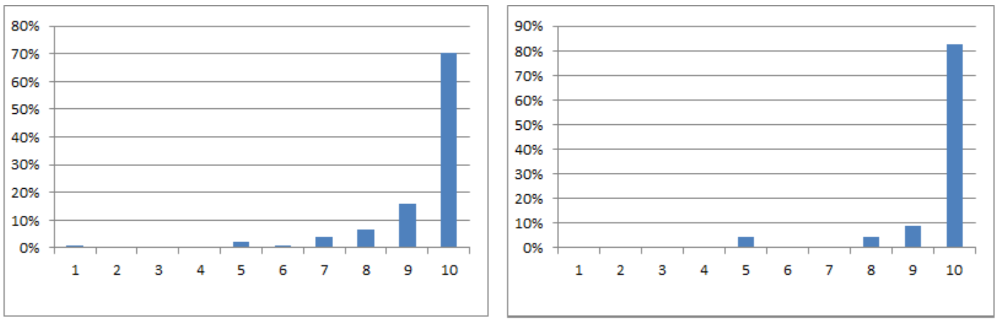

Figure 2: Edit distance split for Documentation and Knowledge Base content

Finally for an end user confidence we did some usability studies on samples of MT where users scored segments on a scale of 1 to 4 where 1 was bad and 4 was good. During our studies we did notice that automated metrics such as BLEU and Edit Distance correlated somewhat with our human usability tests on individual segments but lesser so with a usability test done by a trained linguistic reviewer when looking at the overall project. We put this down to individual strings such as long strings (in excess of 15 words) which caused issues due to writing style.

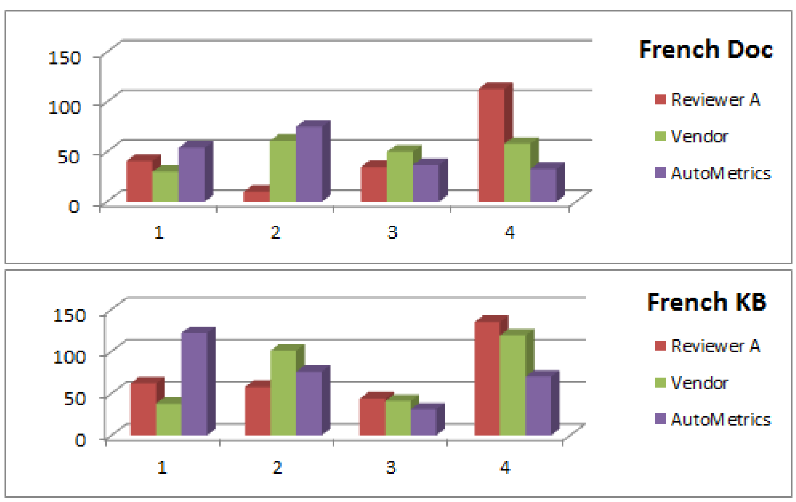

Figure 3: Usability results for Documentation and Knowledge Base content

We have some ability to control standards in our Technical Authoring process, so studies such as Roturier, J. (2004) on Controlled Language (CL) rules effect on MT systems also gave us inspiration for measuring source appropriateness. The idea is that if you have good controlled source authoring style and terminology, then Machine Translation will work better. To understand the nature of this within the company we undertook the task of rewriting some source content to be in a controlled language style. We used this in addition to our normal source content to be Machine Translated for benchmarking in the discovery stages of project. The CL rules we used were based on a limited standardized terminology set and some other basic rules such as sentence length. The two types of content we put into a Controlled Language were standard Technical Documentation; Software Doc/Help and Knowledge Base Articles. It should be noted that by creating new source content and style we must expect some impact on the MT statistics as the non-Controlled Language style is used previously to populate the Translation Memories which in turn are used in training the MT engines.

Using the Edit-Distance, we counted how many segments did not need editing or needed only a low amount of editing and we could see that there was a higher percentage of 100% Match or Fuzzy Match segments that did not need to be Post-Edited with the content rewritten for Controlled Language. This already showed a clear difference between normal authoring and Controlled Language Authoring with regards to the effectiveness of the MT system when displayed across 4 different MT systems. Fig. 4 pertains to the final 4 engines being tested relative to the throughputs recorded in Fig. 1.

	#1	**#2**	**#3**	**#4**
Doc	12%	26%	6%	13%
Doc - CL	17%	30%	13%	19%
Difference	5%	4%	7%	7%
KB	20%	45%	21%	23%
KB - CL	31%	58%	31%	45%
Difference	11%	13%	10%	22%

Figure 4: Percentage of 100% matches for Content v’s Controlled Language Content

It could be said in hindsight that some of our testing was unnecessary. We used a number of other tools that were readily available such as Reading Ease metrics (Flesch-Kincaid and others), measuring the segments and words, average words per segment etc. We wanted full visibility on anything we could measure that may affect output and studied these source metrics on various content types. Despite these measurements not being immediately necessary, they were easy to do, and the lessons learned during this phase do help in the future such as in the ability to notice a high level problem in the authoring process if the numbers move greatly on a particular topic within the Content Management System (CMS).

Corpus Analysis

Source Corpus	Technical Documentation/Help	Knowledge Base Articles
Flesch-Kincaid Reading Ease	51.5	48.2
Grade Levels		
Flesch-Kincaid Grade Level	9.8	9.3
Gunning-Fog Score	11	9.6
Coleman-Liau Index	13.5	13.5
SMOG Index	9.5	8.8
Automated Readability Index	9.6	7.5
Average Grade Level	10.7	9.7
Text Statistics		
Character Count	11,529	11,685
Syllable Count	3,835	4,089
Word Count	2,319	2,345
Sentence Count	153	214
Characters per Word	5	5
Syllables per Word	1.7	1.7
Words per Sentence	15.2	11

Figure 5: Readability Metrics for Documentation/Help and KB Articles

3.1. Make your Enterprise change

A large company does not change processes at the speed of light, it changes slowly and that occasionally makes you actually wonder if it is changing at all. To be the one who tries to turn the enterprise ship can be a daunting task. At the start of discovery in MT it is important to act as an Influencer. This role is to point out areas that could change and the improvements that could be made, find reasons where MT could help and see where people react. Build up buzz around the topic, prove some results and educate your colleagues. These approaches help show what life might be like in the future with a new practice and may demonstrate the potential value of change. Through this movement, followers join your cause and MT will start to go from a topic of conversation through to being involved in projects. People who believe in you and in what you are doing are allies you need to gather in order to make MT a reality. Every conversation potentially helps the cause as many pre-conceptions can exist due to peoples personal experiences with Google Translate and such.

The Enterprise requires due diligence, so every step towards rolling out Post-Editing for productivity should be layered with tests, discussions and some time for people who are not living in the MT or academic world to consume and understand the results. To help in this we have used standard metrics for our organization and we continued this with use of the “Trados Grid” which has an industry understood breakdown of matches in a TM. In Figure 6 we used a GNU license application called KNIME which has some ability for custom workflows of text analytics.

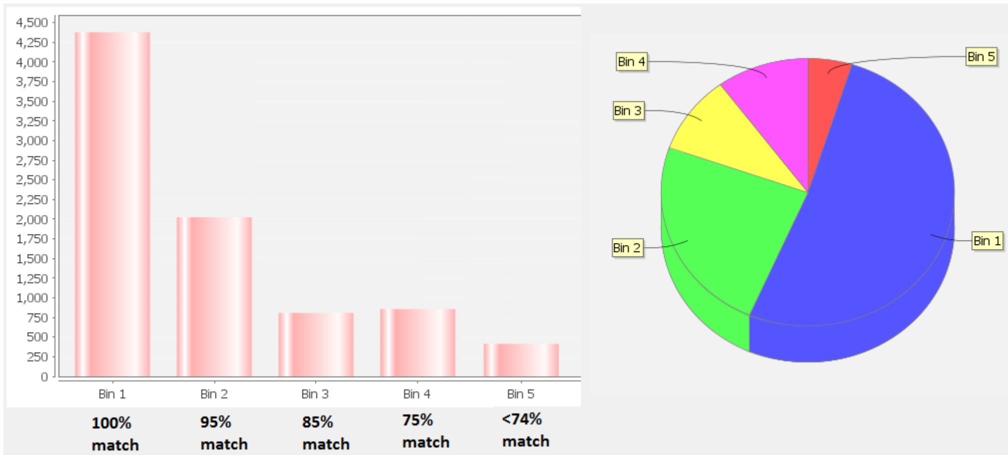

Figure 6: MT Distance matches shown in standard Trados breakdowns using KNIME

It can feel like an uphill battle sometimes to get organizations to change. There is so much to prove to ensure the business case. Luckily there are some very useful resources which can help you prove the theories you preach. The MT suppliers often provide excellent information that can be reused and you should ask them for this if in doubt. Research and white papers can be very useful and many of the MT users meet, collaborate and share their experiences at conferences and online. A newcomer to this area could do well to make friends and ask questions as it can be through these connections that you may build confidence in your own ability to make the best steps forward. Not every approach works for every person or company. Compare, focus and learn the appropriate subjects that you need and ultimately help guide your company towards the right path.

There has to be a need or a gap that can be filled by using MT technology. The first thing that should be done is to identify what the specific business needs are. Having more than one business need gives you a platform to build a proposition to show value and Return on Investment (ROI) in this area.

Business needs for MT in localization are born out of content and publishing. The content gets created and needs to reach an audience. MT can boost the efficiency of that effort through a number of different ways and these are the high level unique selling points for your internal customers that you need to find and understand. Some basic business needs are:

- Increase productivity of translation (Plitt, M. and Masselot, F. 2010).
- Allow on-demand translation for content that normally does not get translated.
- Enable internal users to have access to a larger set of content in their language (Burgett, W., Chang, J., Martin, R. and Yamakawa, Y. 2012).
- To speed up a process of collating sentiment analysis from content.
- Help understand the “gist” of text not available in your own language.
- Enable early versions of localized Documentation or Software.

For the purpose of focus in this paper, we are looking at the productivity of translation as it is an area that can show immediate financial savings. Through increasing productivity many lessons will also be learned and skills gathered necessary for many of the other areas of the value proposition while aiming to save money and time.

3.2. Is MT all that is needed?

This is not only about the MT system as already mentioned earlier in the paper. In some regards it is not even about the MT system itself as the investment in quality of commercial MT systems has been good over recent years, and they continue to get better. The internal workflow is probably the area where most change must happen. This affects a set of items from content writing and curation through how you manage your bilingual content and right down to the end result of publishing and the feedback loop back to the content creation and MT maintenance.

Mentioned previously, Controlled Language Source is probably the most important area to start making changes as it can have a massive impact (Roturier, J. 2006) (Doherty, S. 2012) if it is done right. If you are thinking about rolling out MT in your company, you should start here. Who is writing content that will eventually end up being machine-translated? Is the content good for translation? Does anyone need to change their work practice? After all, in a Globalized company, the source content is most likely only a small percentage of the content distributed to customers around the world. If we can control our language, all the target languages will benefit.

Some basic rules on the content creation side can have a great impact, and consequently, effect on throughput and accuracy in the future. We did some after the fact analysis on Distance per segment and noticed some patterns. The basics that seem to make a big difference are:

- Managed and maintained terminology for authoring – reduction of synonyms
- Basic style rules – keeping all authoring similar
- Reuse of repetitions and phrases in writing
- Source content profiling your authoring into groups (Domains) for MT systems.

Translation process is the next area that needs attention. At the end of the day, the translators are your direct link to your market, and to ensure the best language quality possible and the most accurate message possible, they need to be included in your plans. On paper you may have MT systems with high BLEU scores, but does this become good PEMT in the end? The most important factor towards good quality of PEMT is the translator. In our initial PEMT tests we identified a wide discrepancy in results of translator productivity and quality.

SPANISH	Vendor A	Vendor B	Vendor C
Doc/Help Throughput / day	10688	15280	21360
KB Throughput / day	7464	14088	28392
Quality (LQA Score)	93%	35%	66%
Years experience with Intel Security	5	0	3
Years experience as Translator	7	3	3
Years experience as Post-Editor	1	3	3

Figure 7: Throughput PEMT Spanish with basic Translator Profile info.

After analysis it seemed that the one variable in the process was the individual doing the post-editing. We could not effectively baseline results from one group of post-editors to another with this variable so we needed to reduce or eliminate it and started looking at the concept of Translator Profiling. We would like all translators to translate at the same rate and produce the same quality. This just isn't the case. So there are several parts of a profile that can vary the results, such as individual motivation, or when using freelance or crowdsource models where profiling isn't possible. But as quality is the main requirement for our PEMT there were a number of factors that stood out as a requirement for Translator Profiling for us:

- Experience as a translator is important.
- Post-editing experience is less important (but needs to have some. 2 years is good)
- Age is not necessarily a factor, but (technical) ability to leverage tools might be
- Understanding the content subjects is the most important aspect to reach quality.

What this basically means is that PEMT resources are needed who have spent a good amount of time working on your content so that they understand both your content and your quality expectations. This is evident from Vendor A who has experience on our content and scored high on quality, but lower on throughput. But Vendor B and C did not meet our quality expectations (despite Vendor C having 3 years with our content). The number of years' experience in post-editing is important to throughput with Vendor B and C retaining very high productivity but Vendor A was slower.

The workflow is also an important area to look at. Translation Management Systems (TMS) exist with MT plugged in through APIs. There are other decisions you need to make for your workflow though. Can you trust your translators to not make mistakes such as missing a file for translation? Do you review to ensure no mistakes? Do you allow PEMT segments back into the Translation Memories (TMs) of your main products? Does your TMS edit content before going to the MT system (such as protecting tags or internals)? Do you apply a match penalty on your MT segments and by how much? There is no quick answer for these questions. It is crucial then to understand the nature of the content you want to MT (source) and the nature of the market where you want to publish it (target and quality). Basic localization decisions from normal workflows may need to be rethought when you include MT into the translation strategy.

3.3. Testing MT

Ultimately you will need to test the MT output. Whether you create the MT systems yourself or use a service, or even outsource completely, it is imperative that you run a test. What you want to achieve from this test is a confidence that the standard of quality is high enough on the output for post-editing to happen with extra efficiency.

- Productivity
- Quality (Automatic Estimation)
- Quality (Human Evaluation)

For our tests on productivity we decided that time data was the most important. There are other ways to conduct a test, but at the end of the day if a translator takes less time to post-edit than if they were to translate from scratch, then you are on the right path. Time data is difficult to track, but thankfully over the last few years Computer-Assisted Translation (CAT) tools have evolved to start measuring this (Moran, Lewis & Saam 2014). For our tests we used iOmegaT, which is an adaptation of OmegaT, to gain access to the instrumentation and telemetry on the translator's activities while they post-edit. There are some privacy concerns with this initially; however we found that all translators were happy to be involved once this is part of a test and they had some control over when the feature could be turned off on their console when not in test.

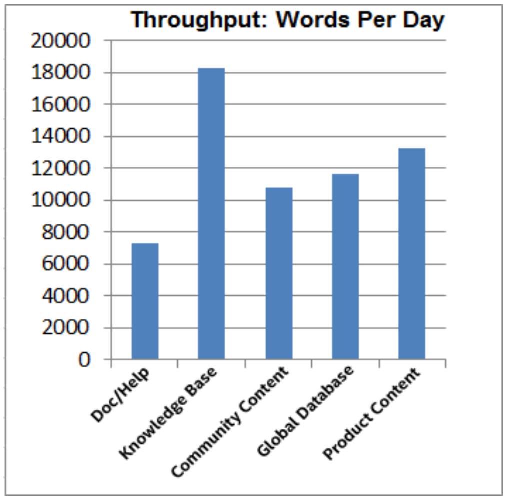

Figure 8: Throughput per type of content showing efficiency compared to others

When we say time data, we basically mean the time it takes to post-edit a segment. It should also include subsequent visits to a segment (not just the first attempt) as it can be a common practice for translators to revisit segments after getting a feel of the document they are translating. Some segments will show erroneous measurements, which is explained when a translator takes a break while having a segment open. We allowed an adequate amount of time for any research the translator may have to do, but we did apply a cutoff to reduce the inclusion of these segments in the test. Timing data can be measured in actual time (ms), but what we used is "words per day" throughput as this is something that most people in our company will understand quickly and easily.

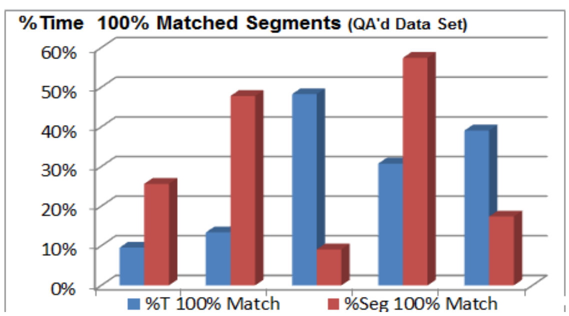

Figure 9: Time spent on 100% match segments from MT

From time data we can learn a lot if we further refine the application of the time data to other metrics such as the brackets for length of segment, the number of repetitions, the number of segments that do not need editing or the number of segments that need a lot of editing (long time segments). We noted that for some content types 100% matches required excessive time to complete when compared to the time spent in other content types. Upon questioning some of these results, the respondents claimed that some 100% match segments require more time to read and understand before they agree that the MT segment is of correct meaning and language quality. These results helped build up a profile of the content for “post-edit ability” and in turn make some recommendations back to the writing teams.

In our discovery tests we sampled 5 different content categories that all loosely communicated within the same domain. While the writing styles and lexical complexities may diverge in their own subdomain, the core subjects are the same. Nevertheless this process was worthwhile as we learned as much about our own internal content as we did about the ability for MT to work with it. You could say that this practice taught us a lesson about Content Profiling before pushing a content type through an MT workflow.

Moving on to quality, the world of Six Sigma says that “Quality is what your customer wants”. So before we enter into a linguistic quality test we should keep that in mind. It’s not often that a customer will come to you and tell you what they want, so we use trained domain linguists to conduct a linguistic quality analysis as appropriate and added some segment usability scores and automatic/automated algorithms such as General Text Matching (GTM) or Levenshtein distance to the test matrix.

Language Quality Assurance (LQA) in its traditional form proved to be sufficient in this case for Quality Evaluation (QE), but more advanced error topologies could also help such as TAUS Dynamic Quality Framework. But with so many factors affecting each segment, you must consider these with a soft focus on the overall quality output as some aspects may need to be prioritized or weighted as having an increased effect on the overall output. What that basically means is that LQA parameters need to be aligned with the quality expectation, and this is hard to manage if you are to baseline quality evaluations against something that may be subjective. In the end, you need a number to go by, but you may actually be more interested in the details of the test (accuracy errors, terminology errors, priority or severity of errors etc.) than the overall score achieved by the text.

To balance the linguistic assessment of your MT text with opinions from would be consumers, usability studies can be carried out. We did a number of tests in our company with various native French, Spanish, Chinese and Italian speakers. The test involved going through 100+ segments and scoring them out of 4 (1 for bad and 4 for great). To get a better idea of the diversity in the scoring, we then applied the same scale to LQA and GTM scores (breaking down the percentage brackets into 1 to 4). From the graph below (Figure 3.) we can see the 186 segments in this test more or less correlate to the same sentiment across the 3 types of measurements applied. The results in this case show that the MT for this language is mostly good and there are a minimal amount of truly bad quality segments.

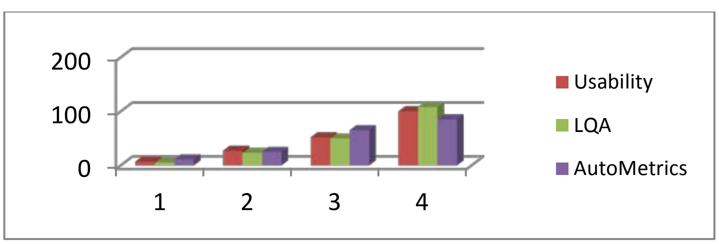

Figure 10: Correlation between 3 quality evaluations (Content Type 1: Documentation)

There is one more thing that we learned while testing our MT and content types. Some content is more prone to error than others when using a static set of Statistical Machine Translation training data. This seems like a reflection on the domain appropriateness, the quality of the MT training and the content writing governance. So we looked at the results and applied "error probabilities" as something to track for the future. The error probability is almost like a predictive metric, but it is tracked at the end of the process so you can learn lessons for the next time. There seems to be a correlation between error probability and "posteditability", which is the intangible measurement of how difficult or easy it will be for a translator to post-edit a segment and ultimately achieve higher productivity while not sacrificing quality. This was seen when a content type and the time it took to post-edit that content type were taken into account while looking at the number of errors. Ultimately this is like a basic type data that one could use when content profiling.

Type	**Segments**	**Error Segments**	**Error probability ratio**
Content 1	152	102	0.671052632
Content 2	23	12	0.52173913
Content 3	11	10	0.909090909
Content 4	94	31	0.329787234
Content 5	23	19	0.826086957

Figure 11: Error Probability rates per content type

3.4. The workflow, the whole workflow, and nothing but the workflow

In localization we may be guilty of looking at the workflow as being the point where we push our content for translation into a TMS. This may be the old way of doing things, but plugging in MT now means we need to know a little bit more about both ends of the workflow. People often talk about "moving upstream", and this means being more involved with the individuals who write content in your company. They are also part of your company's workflow even if you don't control their part. If they can understand the value of making changes to their process, then they will be able to do that better for you. Similarly you need to deliver the multilingual content to your audience, but the audiences are not going to move upstream and tell you what they want, so it's up to the localizers of the world to better understand product promotion, business compliance, marketing and sales.

Moving upstream as a localizer means that you need to get the message across to anyone who writes content, that if they did it in a slightly different way, the rewards to the company could be a great advantage. The basics of content optimization from authors are Terminology, Style, Reuse and Governance. If many authors in the same company can write in a similar way and reuse the same terminology and phrasing, MT will work well for their content in addition to other benefits.

Working with translators can be very useful and they are not that far downstream as they work with the localization teams every day. Their importance comes from relying on them for quality translations but also because they are possibly the first people to read your documentation outside the company. So, if the translator sees something wrong with the content and you have an ability to track their comments, you have a system to create continuous improvement loops on both content accuracy and style of writing. Further downstream again you have the deployed content, and by tracking who uses your product and what they are clicking on, you can further make strategic decisions on the usage of MT and Post-Editing.

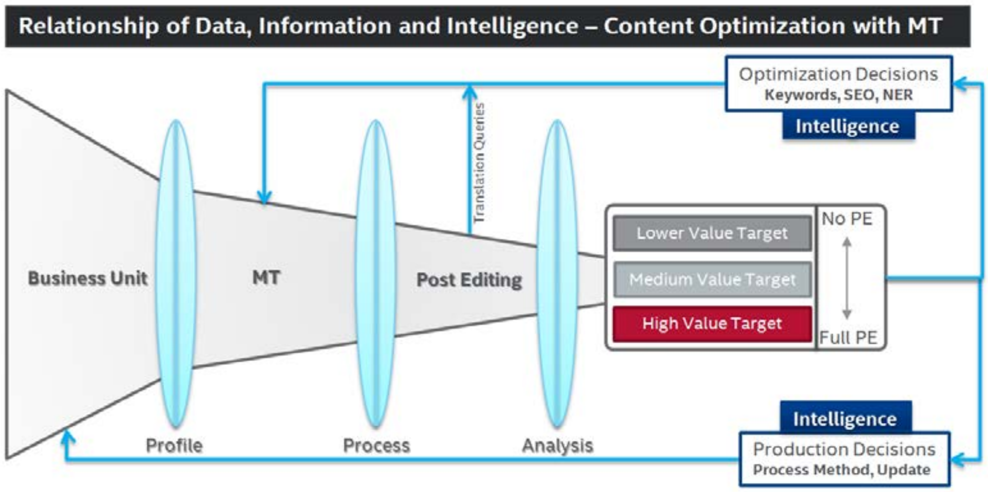

Figure 12: Relationship between upstream and downstream to help MT

So the modern translation workflow, when you include MT, can be improved by working on your source. And the output and consumption can give you great insight into what is working and not working, and needs to be brought back in so you can create a continuous improvement loop.

4. Conclusions

We used methods in this work that draw from 2 basic principles that Controlled Language will help MT output and PEMT can increase throughput in the localization workflow. Our results show that there were significant advantages in using Post-Editing in this case and due to our situation where we could influence the writing standards with our Technical Publications teams these were good starting points for us. Furthermore we understood that some source corpus would be more prone to error when compared to others. While we haven't fully investigated why these differences are yet, we at least have confidence in the texts that do MT well and a basic understanding of the importance of Content Profiling.

The technology is evolving in this area in both the back end of the MT systems and also the new front end Post-Editing Environments being made available. The access to systems such as iOmegaT gave us confidence in the measurements and results when compared to CAT tool agnostic systems such as TAUS DQF for MT QE alone and this real data allowed us to digest a lot of the sales jargon from various suppliers of both MT and Post-Editing services. Having said that TAUS DQF and systems alike do have a place in the process for more sophisticated error topology.

From a higher level in a large corporation, MT can only grow with help from others. We learned that one must spend a lot of time working with the problems of the internal customers while offering the MT solution. ROI must be taken into account a lot at the start so particular focus must be spent on Productivity and Quality Evaluation methods. If the MT project doesn't save money, it's hard to make it grow.

Working with more than one MT supplier can help broaden knowledge quickly: the free or cheap or trial services are ideal to gain insight, learn and build knowledge.

Developers may be needed, either internally or as part of an outsourced partner as there are very few out-of-the-box-solutions and none that will fit all scenarios. But be wary of customizations to TMS and other systems as they can have a costly life length during upgrades.

And finally, once PEMT starts to work, we have the opportunity to look for other ways to use MT in the company. PEMT is a perfect launch pad for the MT program in your company as it is a way to save money and show ROI.

5. Acknowledgements

I would like to thank the translation providers who were part of this project, the Program Managers in the company who helped make these tests possible with their content, the Technical Publications group and the Engineering team. I would like to also thank the various people who helped with multilingual usability studies across the Intel Security site in Cork. Finally I would like to thank Dublin City University and the Center for Next Generation Localization for their talks, expertise and advice.

Glossary

- Machine Translation (MT)
- Statistical Machine Translation (SMT)
- Rules Based Machine Translation (RBMT)
- Post-Editing Machine Translation (PEMT)
- Human Translation (HT)
- Controlled Language (CL)
- Return on Investment (ROI)
- Gist - The substance or general meaning of a speech or text
- Bilingual Evaluation Understudy (BLEU)
- Translation Management Systems (TMS)
- Translation Memories (TMs)
- Computer-Assisted Translation (CAT)
- General Text Matching (GTM)
- Language Quality Assurance (LQA)
- Quality Evaluation (QE)

References

Doherty, S. (2012). Investigating the Effects of Controlled Language on the Reading and Comprehension of Machine Translated Texts: A Mixed-Methods Approach

Roturier, J. (2006). An investigation into the Impact of Controlled English Rules on the Comprehensibility, Usefulness, and Acceptability of Machine Translated Technical Documentation for French and German users.

Burgett, W., Chang, J., Martin, R. and Yamakawa, Y. (2012), Enabling Multilingual Collaboration through Machine Translation. URL http://www.intel.ie/content/dam/www/public/us/en/documents/white-papers/enabling-multilingual-collaboration-through-machine-translation.pdf

Plitt, M. and Masselot, F. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context.

Schmidtke, D. (2005). Microsoft office localization: use of language and translation technology. URL http://www.tm-europe.org/files/resources/TM-Europe2008-Dag-Schmidtke-Microsoft.pdf

O´Brien, S. (2005). Methodologies for measuring the correlations between post-editing effort and machine translatability. Machine Translation,

Claesen, J. Quality Metrics for Machine Translation Output. http://www.yamagata-europe.com/en-gb/blog/item/909/quality-metrics-for-machine-translation-outp

Moran, J. Lewis, D and Saam, C. (2014). Towards desktop-based CAT tool instrumentation.

Plitt, M., and Masselot, F. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. The Prague Bulletin of Mathematical Linguistics, 93(-1). http://doi.org/10.2478/v10108-010-0010-x

Roturier, J. (2004). Assessing a set of Controlled Language rules: Can they improve the performance of commercial Machine Translation systems. In Proceedings of the international conference translating and the computer (Vol. 26). Retrieved from http://www.mt-archive.info/Aslib-2004-Roturier.pdf

Thiel, K., and Berthold, M. (2012). The KNIME Text Processing Feature. Retrieved from http://www.knime.org/files/knime_text_processing_introduction_technical_report_120515.pdf