

Application of Post-Edited Machine Translation in Fashion eCommerce

Kasia Kosmaczewska
TranslateMedia
11 Carteret Street
SW1H 9DJ London, UK
<name>@translatemedia.com

Matt Train
TranslateMedia
11 Carteret Street
SW1H 9DJ London, UK
<name>@translatemedia.com

Abstract

Machine translation (MT) and post-edited machine translation (PEMT) have traditionally been explored primarily in the context of legal and medical content types, where MT results are often easier to predict due to the heavily standardised language structure and unambiguous nature of terminology used. Each content type and domain presents its unique challenges to both MT systems and linguists performing the post-editing tasks. This paper describes how PEMT can be applied in the fashion eCommerce domain, taking a popular British fashion brand – Topshop – as an example. This paper aims to explore different aspects of delivering PEMT to a fashion eCommerce client, the most prominent being linguists' involvement in machine translation-related activities, including their key role in transitioning from human translation to statistical machine translation (SMT), and then from SMT to neural machine translation (NMT). The implications of switching from full human translation to PEMT for the end client and overall learnings made by the language service provider (LSP) during these transitions will be also discussed.

1 Introduction

With machine translation technology going through a period of intense development, the focus of the industry often shifts away from the

human actors without which this technology would not have emerged in the first place.

Taking Topshop as an example, this paper aims to analyse the role of humans involved in the PEMT cycle, namely:

- the client who orders post-editing jobs;
- linguists who handle the post-editing;
- account managers and project managers who oversee the process on the LSP's side.

This is an attempt to describe the impact that the emergence of machine translation technology and related services such as PEMT has had on human actors in the localisation chain.

While machine translation has undisputedly allowed for more automation and increased time-efficiencies in a number of localisation scenarios, it is of utmost importance to evaluate how it has impacted the way fashion eCommerce clients, linguists and LSPs work, in order to be able to continue leveraging this technology adequately in the future.

2 Specific challenges of translating for fashion eCommerce

Providing translations for any domain has its own specific challenges that need to be effectively addressed.

Fashion eCommerce is of growing importance to the global economy as fashion retailers have been quick in grasping the opportunity to grow their businesses internationally, increasingly via online channels.

In 2012, overseas sales accounted for over 13% of total UK online sales. It is predicted that online sales from outside the UK will rise dramatically from circa £4bn generated in 2012 to an estimated

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

£28bn by 2020, accounting for around 40% of total online sales revenue.¹

When going global, fashion retailers often opt for localising their content, and this decision is supported by research by Common Sense Advisory showing that consumers prefer to buy in their own language (Sargent, 2014). One of the most prominent challenges that linguists and LSPs are presented with when working for fashion eCommerce clients is correctly and accurately predicting what would appeal to the target audience in the target market in terms of the tone of voice, vocabulary and the general feel of the target language copy, in a way that satisfies the brand managers in each target location.

Meeting these specific challenges can often mean the difference between the client making its international trading programme a success or a failure. If the target consumer identifies with the way that they are being addressed by a foreign fashion retailer in their native language, then they will be more likely to purchase goods produced by this retailer. On the other hand, if the consumer finds the style and tone of communication in their target language inadequate, it can damage the brand's image in the target market and result in the brand becoming unsuccessful in that location.

When transitioning the service provided to Topshop from full human translation to PEMT, it was of paramount importance not to lose that focus and for the switch between services not to have any detrimental effect on the tone of voice of the localised content.

3 Client profile

Topshop is a global fashion retailer with over 500 shops in 58 countries and, at the time of writing, a considerable eCommerce presence. Topshop has been continuously proactive in communicating with its global consumers using highly targeted, localised content since taking its eCommerce website international in 2011.

Given the international nature of the business, the brand has extensive localisation needs. To address these needs, the LSP has worked with Topshop since 2011, providing translations of brand messages, features, articles, blog posts and product descriptions. The high quality of the translations is of significant importance given that vast majority of these serve to communicate directly with consumers and therefore shape the brand perception in their minds.

3.1 Topshop product descriptions

Topshop's largest requirement in terms of volume is localisation of product descriptions published on the transactional eCommerce website www.topshop.com. With dozens of new products being added to Topshop's website on a daily basis and hundreds of others regularly requiring updates with new specifications, prompt turnaround times have been an important aspect of the localisation cycle for the client.

Apart from needing to put products online as soon as possible, Topshop had a particular yet not unique concern that publishing products on different language websites at different times would disappoint customers. If products become available on the UK website first, customers from other countries are likely to enquire why the same goods are not available on the local version of the site or order goods to be shipped from the UK to their country, which creates operational issues for the retailer and puts a strain on its customer service teams.

To respond to that urgency, since the beginning of the collaboration, Topshop has been using a custom-built integration between their eCommerce platform and the LSP's translation management system to send all products requiring translation automatically, without the client having to manually flag any new items for translation or press a button to make new and updated product descriptions available to the LSP for translation.

Each product description contains between 20 and 100 words and each one of them follows the same structure – it begins with a product name, which is no more than 100 characters long, followed by a long description. The long description comprises of a specification of the product, including washing or handling instructions and garment material composition.

Each batch of product descriptions, containing between 40 and 100 items on average, is turned around and sent back to the client's system within 24 hours so that it can be promptly published on international sites. At the time of writing, this amounts to 1,332 words on average per day, including weekends.

¹ <https://www.statista.com/statistics/284559/e-commerce-sales-of-retailers-in-the-united-kingdom->

[by-foreign-and-domestic/](#)

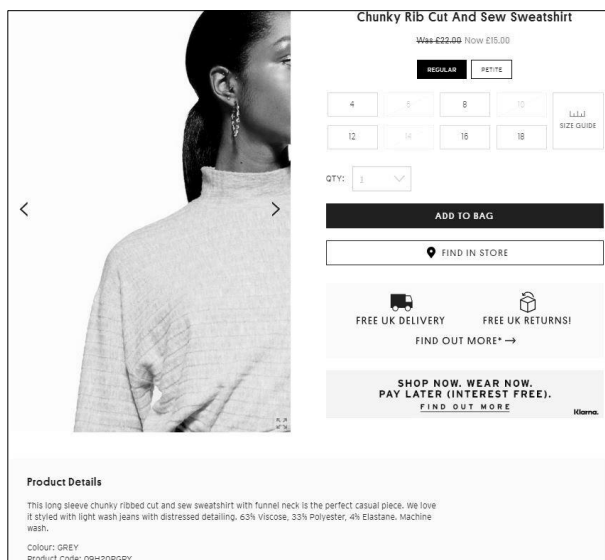


Figure 1. An example of a Topshop product description in English²

4 Switch from full human translation to post-edited machine translation

Five years after Topshop started collaborating with the LSP, it became apparent that the translation memories had grown to over 2m words and the translation memory savings slowly started to plateau.

Due to the nature of the content – well-structured, often repetitive product descriptions – and the language combination in question (English into German), switching the service provided from human translation to post-edited machine translation seemed the optimal way forward to further improve turnaround times, keep the current quality and consistency of the translations and also allow for cost efficiencies on the client's side.

In 2016, the LSP decided to appoint an external machine translation provider to support Topshop's needs. That provider built a custom statistical MT engine based on Moses using publicly available generic data as a baseline and supplementing it with Topshop's translation memory and term bases. The translation memory used to train the first iteration of the model contained approximately 204,641 translation units.³

It was important for the LSP to maintain the positive working relationship with linguists who had been working on Topshop translations over the years. Therefore, the LSP made sure the linguists played a key part in transitioning to PEMT

by involving them in planning, testing, and regular group video conferences to discuss the project.

In the human translation workflow, there were two linguists involved – one translator and one reviser. The LSP wanted to retain the same number of linguists in the PEMT workflow, with the change that the first linguist would act as a post-editor while the second continued as a reviser. This workflow was chosen in order to ensure high quality and consistency of the target language copy, especially in the period when post-editors were still learning how to perform post-editing tasks and finding the right balance between under- and over-editing. Another reason for keeping two linguists in the process was to continue providing a regular stream of jobs for all the linguists who have been involved with Topshop rather than reducing the number of jobs available to each linguist. This was a way for the LSP to show their recognition of the efforts that the linguists have invested in working with Topshop over the years.

In the initial stages of working with statistical MT engine, the LSP had a core team of four post-editors and one reviser working on the post-editing jobs.

Prior to the change, the LSP's internal Topshop team, comprising of an account manager and two project managers, was trained on machine translation technology so that they were able to better assist linguists with any questions and feedback that could arise in relation to post-editing after the go-live date.

4.1 Post-editing with statistical machine translation

During the first phase of implementing PEMT on Topshop product descriptions, which lasted approximately 2 years, the crucial part of the transition was training all involved linguists on how to perform post-editing, including becoming sensitive to the type of errors specific to machine translation systems (Daems, Vandepitte, Hartsuiker and Macken, 2017) and how to address them effectively.

This meant a significant change in the way linguists were used to work for this client (Doherty, 2016). It also meant that they needed to acquire a new set of skills since post-editing presented them with different challenges in comparison to standard translation.

the LSP did not record the exact number of translation units when the initial training was done, the data provided is an approximation.

² <https://www.topshop.com/en/tsuk/product/clothing-427/hoodies-sweats-6864676/chunky-rib-cut-and-sew-sweatshirt-8274864>

³ Since the paper is being written retrospectively and

Since the post-editing effort is directly related to the quality of the machine translation output (Germann, Koehn, 2014), the SMT engine in question was re-trained on a monthly basis with corrections made by linguists during the post-editing stage in order to make sure the quality was improving over time.

The post-editing distance was measured by the external MT provider using the Levenshtein algorithm. Levenshtein distance calculates the minimum number of character edits that are necessary to transform one string into another string (Levenshtein, 1966).⁴

Apart from re-training the statistical engine, in order to reduce the post-editing effort, post-editing rules were also used to automate repetitive post-edits. To create and implement these rules, regular qualitative feedback was provided by the post-editors, who were asked to report on recurring issues that were not resolved by the monthly re-training of the engine.

Throughout the collaboration, Topshop has specified a number of brand rules and requirements that govern the target language copy in German, in the form of a German language style guide. An example of this specific requirement is using proper case in the line name in the German translation, even though it is often fully capitalised in the source. A real-life example of a segment pair demonstrating this requirement can be found below:

Source: ****MATERNITY** Premium White Mom Shorts

Target: ****Weiße** Premium Mom **Umstandsshorts**

This output was achieved through post-editing rules when the LSP was using statistical machine translation.

4.2 Post-editing with neural machine translation

In early 2018, due to the increasing reliability of neural machine translation systems (Srivastava, Shukla, Tiwari, 2018) as well as the increased quality that they are able to return when compared to SMT systems (Volkart, Bouillon, Girletti, 2018), the LSP decided to retire the SMT solution in favour of an NMT system.

The LSP selected an external provider that offers a scalable adaptive neural MT solution. In this provider's approach, the MT output is produced using baseline models trained on generic data but is also adapted in real time to any

similar content stored not only in the existing client-specific translation memory but also any similar translation memories. This approach eliminates the need to train one custom engine for each client or domain, which is very appealing to the LSP in terms of scalability.

This means that MT models are much more agile and flexible – the MT output is adapted to each client's specific style, tone of voice and terminology on the fly. Corrections made by post-editors are sent back to the model upon completion of each job, which eliminates the need for any manual re-training of the model.

At that point in time, the Topshop translation memory contained 282,332 translation units.

The decision to switch from SMT to NMT was made following a period of trials and assessments aiming at verifying whether NMT was going to be a sustainable solution to handle eCommerce fashion content effectively. Comparing results produced by SMT and NMT systems for the same source text is a common evaluation method (Calixto et al., 2017; Castilho et al., 2017). The LSP asked two Topshop post-editors to compare 242 strings translated with the existing SMT engine and the new NMT engine. Each source string was 7 words long on average or 40 characters on average, including spaces. The test set was representative of the live work in terms of content type.

Although each linguist's opinion on individual strings did not always converge, overall both linguists agreed that the new NMT model returned better results than the old SMT model.

In the test, 93 (38%) segments didn't need to be edited when the NMT model was used while only 33 (14%) segments didn't need to be edited when the SMT model was used. These results provided a clear indication for the LSP that NMT was more adequate for the given content type and language combination confirmed that it was time to move onto NMT.

Comparing the output from statistical and neural machine translation engines meant that linguists needed to develop a new set of skills. One of those was sensitivity to machine translation errors which were previously unknown to them, both when they used to provide traditional translation for Topshop and when they did post-editing with SMT engines.

Furthermore, the need for creating post-editing rules significantly decreased when using NMT and the need for monthly engine re-training disappeared completely. As expected, in practice, the

⁴ <http://www.levenshtein.net/>

adaptive neural machine translation framework proved considerably more receptive to corrections made by post-editors, learning immediately from the post-edits, leading to an increase in the number of words being post-edited per hour.

This in turn, also meant that linguists did not need to report on recurring translation errors in the machine translation output as frequently as they did when they worked with SMT. This, as well as the quality of the machine translation improving over time, allowed the amount of time linguists spend on handling Topshop jobs to decrease.

Linguists reported a slight drop in the accuracy of terminology when switching to NMT, however this was conveyed as qualitative feedback and not measured in objective terms. The post-editing effort measurements did not converge with linguists' qualitative feedback in this respect.

4.3 Measuring effort involved in post-editing

Using similar methods to what is described by Federico, Cattelan and Trombetti (2012), since the very beginning of delivering PEMT to Topshop, the LSP has been recording two indicators of the post-editing effort:

- Post-editing speed – the number of words that the linguist is able to review and edit in an hour;
- Post-editing distance – a percentage value indicating the extent to which the raw machine translation output needed to be edited by the post-editor in order to arrive at the desired quality level.

When working with SMT, the post-editing distance was measured by the MT provider. After the switch to NMT, measuring the post-editing effort, including the implementation of the algorithm to calculate the effort, was passed onto the LSP.

Since the first external MT provider the LSP worked with used the Levenshtein distance, this method has also been followed when the LSP started using NMT. In this way, the LSP aimed to ensure that productivity results from before and after the switch could be correctly compared. This measure is mostly being used to unveil trends rather than meant to be absolutely accurate.

To convert the Levenshtein distance into a percentage value, the LSP uses the following formula:

$$(1 - \text{Lev}(a,b) / \text{Max}(|a|, |b|)) * 100$$

Data based on these two indicators – post-editing distance and post-editing speed – has been continuously accompanied with qualitative feedback from the linguists. The feedback has been collected in a number of ways, for example as a list of recurring issues in an online spreadsheet, emails, and also during conference calls between the LSP team and linguists. The subjective feedback covers all aspects of post-editing, including the perceived quality of the machine translation output, the perceived effort invested in post-editing as well as any recurring errors and terminology problems.

4.3.1 Post-editing effort with neural machine translation

While the post-editing distance averaged at close to 38%⁵ when linguists worked with the SMT framework, after the switch to the NMT framework, the post-editing distance dropped dramatically with almost immediate effect to 25% on average. 21.15% is currently the lowest average post-editing distance that has been recorded during a one month period.

Occasionally, particularly in jobs with low word count, the post-editing distance drops below 10%, which had never been reached with SMT on Topshop content.

While the number of words post-edited per hour averaged at 787 with SMT, it rose to 1,000 words and above with NMT.

5 Overall impact

Topshop was one of the LSP's first fashion eCommerce clients to have content localised using PEMT. Not all aspects of the impact of PEMT on Topshop's product descriptions can be objectively or numerically measured⁶. For instance, whether there has been any emotional change towards working on Topshop on linguists' side or client's satisfaction with the post-edited machine translation in comparison to full human translation. Nevertheless, the impact of PEMT has been observed from three different perspectives – the linguists' perspective, the LSP's internal perspective and the end client's perspective.

by asking linguists, the client and project managers involved to provide an account of their perceived experience with PEMT.

⁵ All values in this section have been obtained using the Levenshtein algorithm.

⁶ Although they can be measured in qualitative terms

5.1 Linguists

The team of linguists working on Topshop translations from English into German has not changed significantly since the LSP first started to work with the brand.

This means that the team who had been immersed in the brand tone of voice and all unique requirements at the beginning of cooperation and who used that knowledge when doing full human translation was also able to apply the same knowledge to post-editing machine translation. Skeptical towards machine translation at first, German Topshop linguists were determined to continue providing high quality translations to the end client, regardless of the service through which those are delivered.

After the migration from SMT to NMT, the team of linguists working on Topshop shrank from 4 to 3 post-editors and 1 reviser. This was due to one of the post-editors ending her career as a freelancer.

Migrating to PEMT allowed linguists to acquire new skills and knowledge. They were thoroughly trained on how to effectively work with first SMT and then NMT technology; the knowledge and skills gained are easily transferable to other clients that those linguists work with.

Although there are differences in the post-editing speed of individual linguists, overall all linguists involved in the project work faster when post-editing than when translating, meaning they are now able to complete Topshop assignments quicker than they were able to translate then when they first started working for the brand.

Even though some linguists were rather resistant to the idea of PEMT at the start, their dedication to the brand that they had worked on for several years helped them to eventually overcome the initial reservations.

5.2 LSP's internal team

Applying PEMT to Topshop's product descriptions was an important opportunity for the LSP to verify whether this service can be relevant to fashion eCommerce content in a real-life scenario. In that sense, the success⁷ of the project has led the LSP to offer this service to a number of other fashion eCommerce clients, including some well-known high street clothing brands.

The Topshop example has given project managers and account managers confidence in handling

PEMT projects. Similar to linguists, the internal team learnt what to expect from the raw machine translation in the context of fashion eCommerce content and now understand its benefits as well as its limitations.

One of the most prominent learnings on the part of project managers was adequately assisting linguists with the post-editing task through detailed briefings, done in writing and on conference calls. This also included guiding those linguists who were more resistant to idea of post-editing than others as well as finding optimal ways of handling supporting tasks such as gathering linguist feedback.

5.3 End client

As a result of switching from full human translation to PEMT, the client has been able to enjoy the same quality of the target language copy at a reduced cost.

Since the quality and consistency of the translations have been maintained, the change in service has had no adverse effect on the brand perception in the target market.

Furthermore, it was an opportunity for Topshop employees responsible for ordering translations to also become familiar with MT technology.

At the time of writing, the Topshop translation memory contained 350,776 translation units and continues to grow.

6 Conclusions

This paper aims to demonstrate that PEMT is a viable service option for eCommerce fashion content intended for the German market.

It has transpired that the success of delivering PEMT to Topshop largely depended on ensuring that linguists were well informed and engaged at all stages of the project – that they understood the task well, could rely on project managers to guide them when they had doubts or questions regarding MT, and had a streamlined way of providing feedback. The mutual understanding and good flow of communication have made it possible to efficiently address any issues that arose on the client's or linguists' side in relation to PEMT.

It appears that PEMT can indeed work well for fashion eCommerce content such as product descriptions. Since this service allows for faster turnaround times, while also being available at a lower price point than traditional translation, it opens new localisation opportunities for retailers.

⁷ Defined by client satisfaction and linguist productivity; the latter being directly related to machine translation quality.

References

- Calixto, Iacer, Daniel Stein, Evgeny Matusov, Sheila Castilho, Andy Way, 2017. *Human Evaluation of Multi-modal Neural Machine Translation: a Case Study on E-commerce Listing Tiles*. Proceedings of the 6th Workshop on Vision and Language. Valencia, Spain, 31-37.
- Castilho, Sheila, Joss Moorkens, Federico Gaspari, Rico Sennrich, Vilelmini Sосoni, Yota Georgakopoulou, Pintu Lohar, Andy Way, Antonio Valerio, Barone Miceli Barone, Maria Gialama, 2017. *A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators*. Proceedings of MT Summit XVI, vol.1: Research Track.
- Daems, Joke, Sonia Vandepitte, Robert J. Hartsuiker and Lieve Macken. 2017. *Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort*. *Frontiers in Psychology*.
- Doherty, Stephen. 2016. *The Impact of Translation Technologies on the Process and Product of Translation*. *International Journal of Communication* 10 (2016), University of South New Wales, Australia, 954.
- Federico, Marcello, Alessandro Cattelan and Marco Trombetti. 2012. *Measuring User Productivity in Machine Translation Enhanced Computer Assisted Translation*. MateCat Publications, 3.
- Germann, Ulrich, and Phillip Koehn. 2014. *The Impact of Machine Translation Quality on Human Post-editing*. Workshop on Humans and Computer-assisted Translation, Gothenburg, Sweden, 38–46.
- Levenshtein, Vladimir I. 1965. *Binary codes capable of correcting deletions, insertions, and reversals*. *Soviet Physics Doklady* 10(8), 707–710.
- Sargent, Benjamin B. 2014. *Ten Tricks for Improving Global CX*. Common Sense Advisory, Inc. Cambridge, Massachusetts, United States of America, 1-3.
- Srivastava, Siddhant, Anupam Shukla, Ritu Tiwari. 2018. *Machine Translation: From Statistical to Modern Deep-learning Practices*. *Computing Research Repository*, vol. 1812.04238.
- Volkart, Lise, Pierrette Bouillon, Sabrina Girletti. 2018. *Statistical vs. Neural Machine Translation: A Comparison of MTH and DeepL at Swiss Post's Language*. Proceedings of the 40th Conference Translating and the Computer. London, United-Kingdom.