WHITE PAPER

Neural Machine Translation Reaches Human Parity

WordSynk
Machine Translation

thebigword

Disclamer

The information presented by thebigword Group on
this white paper is for general informational purposes
only. All information on the paper is provided in good
faith, however we make no representation or warranty
of any kind, express or implied, regarding the accuracy,
adequacy, validity, reliability, availability or completeness
of any information on the paper.

Under no circumstance shall we have any liability to you
for any loss or damage of any kind incurred as a result
of the use of this paper or reliance on any information
provided on the paper. Your use of the paper and your
reliance on any information on the paper is solely at your
own risk.

Contents

INTRODUCTION

1. This paper

In this paper you will learn how you can use machine translation to reduce your cost by over 90% and receive instant access to high-quality translations.

Now more than ever, Machine Translation plays one of the most important factors in language localization. So, what is it?

Machine Translation provides customers with a service that allows them to translate digital content automatically from one language to another language. The tools within this can be customed to meet client requirements, including linguistic assets and basic preferences.

Today, we are seeing Machine Translation become more and more common due to advances in Neural machine learning algorithms, which work as a technological alternative to the human brain.

The latest development in WordSynk utilise Artificial Intelligence and a machine learning algorithm to provide clients with an automated analysis of the best workflow for a translation project. First time, every time.

We are adapting the way our WordSynk platform works by using these new developments to enhance the MT experience for our clients. Our Machine Translating with post-editing can save time, money and receive content nearly the same level of quality,

These developments in technology are seeing Machine Translation being used in a wide number of industries where there are lots of data – such as manufacturing – and by public sector organizations such as the Police and Government.

This paper will detail how we are using Machine Translation, how it works for our clients, statistics and example case studies of when Machine Translation has saved clients time and money.

2. How we are using Neural Machine Translation

Advances in technology means that Machine Translation is getting better each and every day. At thebigword, our team our constantly adapting the way our systems work, using new developments to enhance the MT experience for our customers.

Our WordSynk platform makes continuous localization easier and smarter by automating manual tasks and learning from preferences to help save time.

Additional features such as Patented Computer Assisted Translation (CAT) tools restrict linguists to secure content and connectors which allows WordSynk to connect to third-party tools improving Machine translation still further.

2.1 Different type of Machine Translations

These are the different type of Machine Translation Technology available on the market.

Augmented MT
Highly tuned engines and hybrid workflows ensure the best quality document, and additional post-editing can create content comparable to full human translation.

Neural Machine Translation (NMT)
An advanced machine translation technology, utilising artificial intelligence and machine-based neural networking. With the ability to provide custom NMT engines, we provide the highest quality output.

Neural Machine Translation with Human Post Editting

We offer a variety of different post edit activities and customise our post-editing workflows for each client. MTPE solutions focus on creating lasting content that is cost-driven and is suitable for high-volume web content.

Instant Raw Machine Translation

We offer a variety of different post edit activities and customise our post-editing workflows for each client. MTPE solutions focus on creating lasting content that is cost-driven and is suitable for high-volume web content.

3. How does NMT work for clients?

For clients, Machine Translation is a simple process when using our WordSynk platform. The diagram below shows the steps, including the additional Human Post Edit and a QA Check which services which are optional.

| Customer Submits Content | → | Machine Translation | → | Human Post Edit (Optional) | → | QA Check (Optional) | → | Complete Delivery |

4. Clients using NMT

Human Translations

Result: Very Good

Overall, the quality of the translation is good, but it does not always read well. The translation is sometimes too close to the source and not always accurate.

Machine Translation Post Edit

Result: Very Good

The quality of this translation is good, and the writing style is well suited to the document. If I had to compare the 2 translations, I would say that this one is better in terms of style. The translation is easier to read and contains fewer errors than the other one.

Our clients are also gaining more confidence with Machine Translation due to its cost effectiveness and time reducing benefits. In addition to the accuracy of the post-editing option is making MTPE a great alternative to traditional Human Translation.

On a recent quality assurance test between Human Translation and Machine Translation with Post Edit for one of our clients, Honeywell Security, the external proofreader provided this feedback when comparing Human and Machine Translation.

CASE
STUDIES

xerox **Case** Study

5.1   Challenge

Xerox's Corporate Marketing team are responsible for creating and localizing Sales Enablement Collateral (including both brochures and specification sheets) for Xerox products and offerings. They translate their content in 30 languages and spend approximately $500,000 per year on localization.

They approached thebigword to see if it would be possible to move their content to a Machine Translation Post Edit (MTPE) model with a focus on reducing translation spend. They wanted to save more money per year by using this model, as well as reduce the time taken to localize their content.

5.2   Solution

thebigword carried out an extensive implementation process, which initially identified which Machine Translation (MT) Engines were most suitable for the language pairs requested by Xerox.

From here, the next stage was to check that there were over 10,000 translation units in the Translation Memories for all language pairs. At this stage, some languages were excluded as the Translation Memories were not large enough to train the engines. Given the nature of the content, proceeding with an un-trained engine was not deemed appropriate.

A test pilot was created using two of the most frequently requested language pairs as chosen by the client. The languages chosen were Russian, which is a more challenging language in terms of MT quality, and Brazilian Portuguese which typically produces a higher MT output.

Using Artificial Intelligence and the WordSynk platform, the engines were trained using Translation Memory and glossaries and then the quality was tested. The first test was completed with the Bleu score.

What is the Bleu Score?

The Bleu scores uses automated algorithms to determine whether the trained engine is producing a higher output than the standard engine. When the test was completed with the Bleu score, the results were positive for both languages - so the next stage was the Translation Automation User Society (TAUS) method.

What is the TAUS Method?

The TAUS method performs a test using linguists to assess the adequacy and fluency of MT translated strings. When this was tested with both languages, this received a positive outcome again with both categories scoring 3-4, which indicated the process could move forward to the next stage of testing.

The next stage was a pilot project, where thebigword ran human translation and proofing in parallel with a MTPE and proofing workflow on the same project

When both human translation and MTPE files were ready we sent to an independent QA checker. We did not tell the QA checker that one version was human translation and one was MTPE, we just asked them to assess the two translations.

5.3   Results

Once we had completed the implementation process with the initial two languages requested by the client, we ran the same process for various other languages requested.

The results back from the pilot testing showed positive feedback. In most cases the MTPE result was as good as the human result and in some cases the MTPE provided higher quality results than the human translation – especially with French, Spanish and LA Spanish.

The table below shows the full results from the pilot testing:

Language	Human Result	MTPE Result
Russian	Very Good	Good
Brazilian Portuguese	Very Good	Good
German	Good	Good
French Canadian	Satisfactory	Satisfactory
French	Satisfactory	Good
Italian	Good	Good
Spanish	Satisfactory	Good
LA Spanish	Satisfactory	Good

Following the results of the pilot project, Xerox has adopted MTPE plus proofing into their translation workflow, which is estimated to save the client over 25% compared to a traditional human translation workflow with proofreading.

Xerox is continually looking to introduce the latest innovation and technology to streamline efficiencies within our own business and also across our customer accounts. Working with our dedicated Account Management Team and Machine Translation Experts at thebigword we have refactored our workflow.

Integrated Neural Machine Translation (supported by AI) has allowed us to drive down costs, improve speed to market and maintain language quality. We are starting to see the benefits of this approach after a few weeks of launch and expect the machines to continually improve/learn as more flows through them providing us with continued savings over time.

TBW proved that they were able to meet our technological expectations within a timely manner and that they were willing to support the partnership with ongoing value add through a Technology Road Map.

Rickin Hindocha
Xerox Content and Marketing

METROPOLITAN POLICE **Case** Study

6.1   Challenge

Police services investigating large-scale, international, organised criminal groups often gather vast tranches of foreign language material during the course of operations. The volumes of material recovered are growing exponentially against a background of diminishing resources and collapsing timeframes. The Metropolitan Police Service (MPS) have highlighted how the deployment of an efficient MT workflow can be pivotal in reducing costs, saving time and delivering a more robust product to meet disclosure obligations.

Officers from the MPS Modern Slavery & Child Exploitation, Central Specialist Crime Command were working on a case where they required over three millions rows of chat data recovered from twenty digital devices seized from suspects translating from Mandarin and Italian to English across a variety of language settings. The expediency of the process was vital in order to manage risk, identify potential victims of trafficking and highlight key strands of evidence.

The reliance on conventional / human translation is resource intensive; both in respect of the cost of interpreters and the additional time required to manage their workflow. Both factors impact negatively on the progress of the investigation and the public purse. Following this, the MPS approached thebigword for a more expedient and efficient alternative.

6.2   Solution

Keen to embrace more innovative solutions to manage large volumes of foreign language material and drive the brevity of investigations, the MPS commissioned a trial to determine the viability of employing MT in this investigation. Based on the samples provided, thebigword successfully trialled a number of options to identify varying accuracy rates, security / data-handling options and cost implications.

Fully sighted on specific issues ranging from data security protocols to continuity of exhibits, the Officers secured the support of senior managers to sponsor the project.

After assessing the files and seeing the full scope of the project, thebigword were committed to delivering the finished project in just 12 working days.

6.3   Results

The files that were submitted contained in excess of 3 million lines of data requiring translation, which were delivered to the client within the 12 working day commitment. Following the translation thebigword shared a quality summary, which showed the quality scores, information, and data security.

Using Machine Translation, the MPS secured these outcomes with minimal abstraction of time or effort enabling Officers to focus on driving the investigation forward. Both time and cost savings are considerable with the additional benefit of freeing interpreters to focus on the translation of evidential product identified following the review process.

The MPS were happy with the results that they got, and our MT solution suited their needs. Though the main reason for using the solution was to save time, they were able to save cost too.

> We must seek opportunities to make best use of emerging technologies, manage our limited resources efficiently and work to maintain public confidence in our investigations. Maintaining absolute control over the process of reviewing non-evidential material recovered during the course of an investigation is of equal importance to the weight of evidence.
>
> Utilising this technology gave our team the freedom to focus on building a robust and conclusive investigation, manage risk and safeguard vulnerable victims of Modern Slavery.

DS Peter Eddy
The Metropolitian Police Service

7. Machine Translation Statistics

7.1   Cost and Price

Machine translation is nearly instantaneous and costs only pennies per word, meaning significant savings are available.

Most commonly however, raw machine translation will require one or two rounds of post-editing by a professional human translator to ensure human-level quality and fluency of the final translation.

In our experience, machine translation plus post-editing can produce a professional result at about 70% of the time and cost of a full human translation.

We have over 4,000 professional translators performing post-editing work. This allows us to accept hundreds of thousands of source words for post-editing and on average, turnaround time will be about 75% of the turnaround time for full human translation.

70% time and cost saving when using machine translation with post-editing

75% turnaround time with MTPE compared to full human translation

7.2   Benefits of using Machine Translation

Cost effective for high-volume projects

Customisation for improved accuracy, including MT Post Editing

Fast turnaround times

WordSynk
Machine
Translation

Collaborating with a team of in-house MT expert engineers

All file types accepted and returned ready to use

Quality output scoring and reporting, ensuring project performance and continued improvement

CONCLUSION

8.1   What does the future hold for MT?

Features such as post-editing allow humans and machines to work in unity, meaning more demand for Machine Translation will not only help linguists, but help innovate the future of the language industry.

Depending on various factors such as the type of content, urgency of content and accuracy of content. Machine Translation and Post-Edit are giving end users far more choice.

With the right provider, like thebigword, Machine Translation is faster, cheaper. Our advances in technology such as the WordSynk AI Recommendation workflow ensure the best project output first time, every time.

8.2   How our other clients can adopt MTPE

Human reviewers produce far higher volumes of content, at a fraction of the cost and nearly the same level of quality. Our Machine Translation with post-editing can save clients at least one half of the price of traditional human translation.

WordSynk AI uses historic data and logic for its decision-making processes. As it is a data driven workflow, it is not influenced by human bias. The Artificial Intelligence learns from its performance and constantly evolves, the more data is added.

WordSynk will automatically advise you which workflow should be used and then automatically build the workflow in the background, making the process as simple as possible for you.

thebigword is constantly striving to innovate and incorporate developments in technology to ensure that our Machine Translation solution is the best for our clients.

For more information about our translation solutions, please contact info@thebigword.com. We are glad to provide a demonstration to your project stakeholders of how to use Machine Translation on the WordSynk platform.

thebigword

thebigword Group

Link Up House, Ring Road, Leeds, LS12 6AB

0870 748 8000

info@thebigword.com

www.thebigword.com