# Case Study:

## Integrating Machine Translation and Risk Prediction to Achieve Cost Savings

GlobalDoc®
WORLDWIDE TRANSLATION SERVICES

Enterprise content teams requiring high-quality translation services have traditionally insisted that agencies translate from scratch — no post-editing from machine translation.

GlobalDoc ([www.globaldoc.com](www.globaldoc.com)) has translated for teams at enterprises like IBM, Tenneco, Toshiba, Xerox and other high-profile clients for decades — clients who have shared their concerns and requirements with GlobalDoc about the cautious introduction of machine translation technologies within the translation process that eliminate or reduce human translator involvement. At the same time, the massive growth in content and related costs to support languages in established and emerging markets provides constant motivation to use advances in technology, and to automate more.



GlobalDoc, Inc. was founded in 1993 to translate information for Fortune 500 marketing and communications teams. The founder-led company grew from Atlanta to Europe and Asia as it developed its own cloud-based translation management system, LangXpert® to automate ordering, translation process management, delivery, and review in any language for constantly evolving content types.

## CHALLENGES

As a technology-centric language service provider to companies in a wide range of industries, GlobalDoc has followed innovations in machine translation over the decades, but could not confidently recommend a machine-driven translation solution to clients that could both achieve client quality demands while also saving substantial sums of money.

### High-quality enterprise content

Consumer-focused translations apps, powered by off-the-shelf translation APIs, do not work well on enterprise content. Important or sensitive content such as corporate communications, marketing materials, multimedia/web information, and technical manuals often need a human touch. Professional translators spend significant effort post-editing machine translations to maintain quality within traditional translation workflows. Client reviewers in-country, who have jobs unrelated to translation, are often burdened with verifying their company-specific information within translated outputs. All of these elements can make merging machine translation strategies within sophisticated translation processes that support multiple language requirements simultaneously very challenging.

### Effort estimation

Machine translation quality can vary wildly. This makes estimating scope-of-work and costs more difficult when having to analyze data without the benefits of upfront scoring. GlobalDoc understands that clients rely on the company to maintain quality while reducing costs. Increases in content volume across multiple languages and end-use platforms can dramatically increase the cost and time spent on providing translated content. The LangXpert + ModelFront process allows GlobalDoc to accurately predict costs and time elements on estimates presented to clients, while also providing reliable data from which to estimate future cost and time requirements over time.

## Fragmentation

The total volume of translation orders can be very fragmented and varied across clients, business units within clients, and project elements, to include multiple file types across content and requiring consideration for many types of client audiences. But for integrating machine translation, proper setup — starting with know-how about machine translation providers, customization and quality control — is required for each client, project, and language pair. This can be a daunting, ever-changing task.

## Enter COVID-19 into the equation

As the COVID-19 pandemic hit societies around the world in Q1 2020, enterprises felt their businesses and their own operations severely disrupted. They also faced pressure to quickly disseminate critical emergency information in many languages and to launch new businesses. Employees within GlobalDoc's client base were being displaced to working from home—separated from the resources they were accustomed to within their daily routines.

Efficient, quick, and cost effective ways to disseminate information to global workforces suddenly became vitally important to companies — and many turned to GlobalDoc for assistance. Translators were also overwhelmed, both in their availability and in dealing with their own personal circumstances as a result of COVID-19. GlobalDoc knew it was time to look for creative and innovative ways to help clients fulfill these unprecedented requirements to communicate across the globe. And given rising financial pressures and necessary budget changes, doing so in a fast and economical manner.

> "We were already aggressively looking for ways to offer more automation and cost efficiencies to our clients, and the pandemic accelerated the absolute necessity to succeed in doing this seemingly overnight."
>
> **— Michael Cooper**
> Founder and CEO
> GlobalDoc, Inc.

## A NOVEL APPROACH

### LangXpert + ModelFront

Machine translation quality estimation is a topic of open research inside technology companies like Amazon, Facebook, Google, IBM, and Microsoft.

> *Quality estimation: Automatic methods for estimating the quality of neural machine translation output at run-time, without relying on reference translations[1]*

Deep-learning approaches, based on massive multilingual language models, are gaining ground for instantly predicting segment-level quality metrics within sentences, nearly mimicking a human post-editing effort. This approach provides data to consider aggregate metrics, instant document analysis, and overall project-level evaluation to assist GlobalDoc in quickly analyzing available machine translation quality options across all major platforms — in any language.[2]

After reading about ModelFront in an industry publication, GlobalDoc CEO Michael Cooper assessed the landscape and found ModelFront the clear front-runner. The firm's translation risk prediction API and console made it the leading provider of production-strength solutions for quality estimation and evaluation. The ModelFront team, led by CEO Adam Bittlingmayer quickly grasped the direction GlobalDoc was headed—and light bulbs started going off quickly in the heads of both company CEOs and their respective teams. Within weeks, and during the chaos of the pandemic, the two companies were able to enter into a mutually beneficial exciting partnership, without hesitation.

## Partnering with ModelFront

GlobalDoc and ModelFront have [partnered](#) to integrate ModelFront technology into GlobalDoc's translation management system, LangXpert — to openly and transparently share know-how on use cases and translation technology with clients, something many companies in the translation industry have found difficult to offer in the past. GlobalDoc puts client interests ahead of their own balance sheets to provide companies with a much-needed approach to saving money, while maintaining the quality standards that are so important to their worldwide operations.

With GlobalDoc's guidance, ModelFront tunes the accuracy for GlobalDoc's diverse client use cases, develops support for the required project and document formats, and integrates customizable machine translation from the most suitable providers. **Human translators are still involved, but in a much more focused manner whereby they are provided with scored information by segment to concentrate on.** Combining new elements of scoring technologies while maintain high levels of project control against multiple language requirements simultaneously within the LangXpert software allows GlobalDoc to offer clients significant savings — both in cost and time — while maintaining extreme quality control standards within the translated content.

As partners, we now present the solution as an option to GlobalDoc clients, who are able to preview real results behind the scenes — actual machine translation and risk prediction output — on recent projects before they choose the full post-editing option for future projects. Overwhelmingly, clients are adopting this new approach within their corporate processes for disseminating information globally.

## How it works

Let's look at how it worked on a real project from Q3 2020 where the client, a Fortune 500 company in the automotive space, selected the full post-editing LangXpert + ModelFront option.

> **10 instruction manuals as Adobe InDesign® documents (IDML format)**
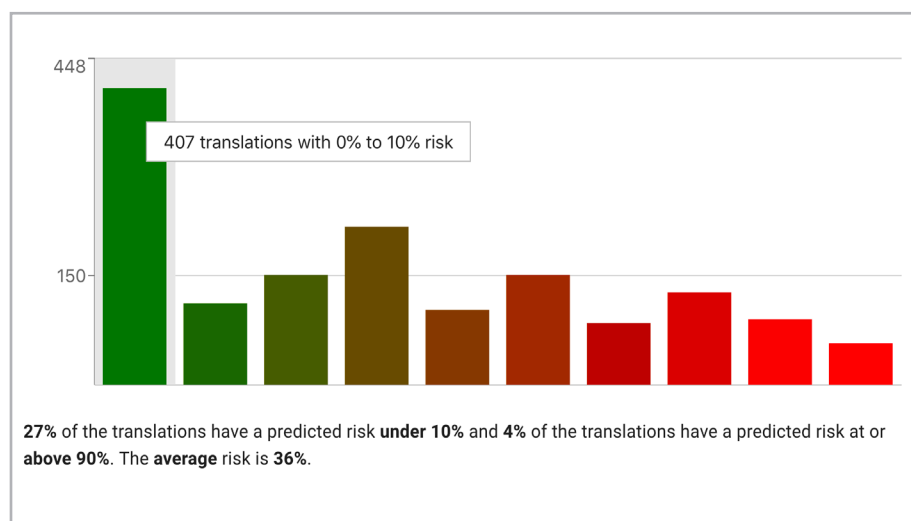>
> 2,595 segments (14,181 words)
> English (United States) to German (Germany)
> 1,095 exact translation memory (TM) matches and
> 1,500 new segments

GlobalDoc's LangXpert workflow, accessed by secure client login 24/7/365, parses the documents into segments — meaningful units of text like titles or sentences — and invokes ModelFront for custom machine translation and translation risk prediction on the 1,500 new segments for which there was no exact translation memory match from previous projects. The machine translation is actively customized over time based on the dynamic translation memory GlobalDoc maintains.

The distribution of the segment-level risks for this project is shown in the ModelFront console:



448

407 translations with 0% to 10% risk

150

**27%** of the translations have a predicted risk **under 10%** and **4%** of the translations have a predicted risk at or **above 90%**. The **average** risk is **36%**.

For this project, many of the segments have very low predicted risk — the estimated probability that the segment will need to be post-edited, even by one character. The system factors in not just the quality of the machine translation, but also the inherent difficulty and quality of the source content. The aggregate score is length-weighted to account for actual post-editing effort.

The dashboard also has a preview where the segments are sorted by predicted risk and labeled by error type, so the project manager can drill down into the actual text in a targeted way.

The distribution of the segment-level risks for this project is shown in the ModelFront console:



```
EN   <g id="5840"><mrk mtype="seg" mid="1906">Spring</mrk></g>
DE   <g id="5840"><mrk mtype="seg" mid="1906">Frühling</mrk></g>                100%
```

As you can see above, The preview surfaces a high-risk segments in the ModelFront console without context, the English word *Spring* is ambiguous — it could refer to the season, a jump, a hardware part or a product or brand. Here it is machine-translated to the season, and the human translator will understand the full context and correct it to *Feder* — a hardware part.

With all this information, the GlobalDoc project manager knows that the professional human translator will be able to review and approve many segments — a significant improvement in speed while maintaining quality control. These efficiencies are passed on to the client in the form of actual, tangible cost savings reflected on client invoices:

## The upfront quoted cost to the client was reduced from $1,697 to $1,102 for additional savings of 35%.



Efficiencies from machine translation

Perfect translation memory matches

Final cost
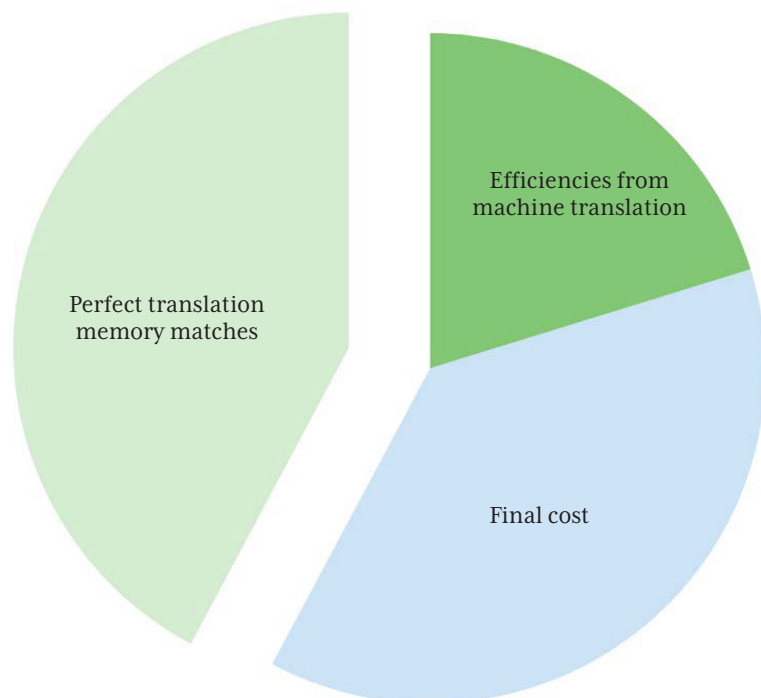
GlobalDoc provided a transparent quote both to the client and to the translators that reflected the actual difficulty tier before the project started. After the project was delivered with the final post-edited translations, the post-editing data is used to evaluate the risk prediction system and continuously improve its accuracy for future projects.

In addition, GlobalDoc updates LangXpert translation memory information to provide a base for improved quality and scoring for future client-specific requirements. The combination of technologies involving professional application of translation memory and ModelFront scoring capabilities should over time provide increased substantial benefits.

The LangXpert + ModelFront strategic approach has since been successfully applied by GlobalDoc to a wide range of document types, clients, and language pairs over Q3 and Q4 2020. The partnership enthusiastically continues with more development activities and process enhancements planned for 2021.

**"The feedback from both our clients and our translators is consistently positive."**

**— David Jett**

Vice President - International Operations
GlobalDoc, Inc.

## SUCCESS FACTORS

The success of integrating enterprise-level content management, machine translation, translation management, and risk prediction technology depends on multiple criteria:

### High-quality source and translated content requirements

GlobalDoc clients require excellent quality, and likewise the company is focused on high client satisfaction. Because clients generally provide their original content in excellent quality and in a style consistent with their previous projects, customized machine translation performs better. It's also important to segment information in a way that's optimal for machine translation. The higher the quality of the source content, the higher the quality and more successful the LangXpert + ModelFront machine translation approach will be.

### Professionally managed and meticulously applied translation memory strategies

GlobalDoc has maintained translation memories for clients over decades. Every word a client sends to GlobalDoc is analyzed within the LangXpert workflow — a project cannot skip this important step, both on the front and back ends of the process (from project inception to delivery). Customization of machine translation mainly depends on the size and quality of the client's translation memories. ModelFront scoring is enhanced by applying translation memory strategies as well.

### Sensible, flexible, realistic, and reasonable approach to automation

GlobalDoc and ModelFront work with clients to educate stakeholders to the fact that not all content is a good fit for machine translation. While some translation agencies and machine translation providers try to make everything work in all languages with respect to automation, GlobalDoc and ModelFront recognize that the human element to a professional translation process requiring quality output is still required.

However, integration of GlobalDoc's translation automation process much less expensive than adoption of other solutions within an enterprise environment, with the introduction of scoring savings an added bonus to automation of the cost saving workflow. Machine translation accuracy depends on the document subject matter, the language pairs and data available for customization. Different machine translation options have different strengths and weaknesses. The LangXpert + ModelFront

**1**

**2**

**3**

process analyzes all of these so clients do not have to understand the intricacies of the process.

Encouragingly, the cost savings, features, languages, and local support of machine translation service providers is constantly expanding.[3] GlobalDoc's partnership with ModelFront has proven that even during the most challenging of times – historic in fact—that innovative, open-minded approaches to humans solving issues by smartly applying technology can help move industries forward positively.

From the overall project to the word-level, LangXpert + ModelFront's cutting-edge technology provides valuable information to human experts who make the final decisions, and who ultimately maintain the final quality of the traditional translation workflow in support of major global content initiatives.

## LOOKING FORWARD

GlobalDoc and ModelFront are actively exploring ways to automate the process of setting up custom machine translation and custom risk prediction for every client and language pair. The companies are actively applying risk prediction and scoring to more use cases, like final review/validation of translated information.

Please contact GlobalDoc for assistance and more information.

### Email

info@globaldoc.com

### Offices

**GlobalDoc, Inc.**
10 Glenlake Parkway
Suite 130
Atlanta, Georgia 30328 | USA
Toll-Free: 800-241-8646
Telephone: 404-350-6740

### References

1. Quality Estimation Task, Fifth Conference on Machine Translation, EMNLP 2020

2. Seven Machine Translation Trends in 2020, Maxim Khalilov, TAUS

3. An overview of the features and limitations of the major machine translation APIs, ModelFront