



# The Power Couple: Machine Translation & eDiscovery

### Introduction

Often, there are certain translation applications where attempting to perform such projects without [Machine Translation \(MT\)](#) will end up a mess. Typically, this happens when the application must deal with a combination of factors including those listed below:

- When there is a huge volume of source content that can't be translated in a specific time frame without MT.
- When the content needs to be translated within a quick turnaround time for it to provide value to the consumers.
- When user tolerance for lower quality translations in the early stages of information review is unacceptable.
- When the highest priority content has to be identified from a huge volume of indistinguishable content to extract information and document triage. This process, in turn, allows for superior quality human translation.
- During translation cost prohibitions are applied.

Many of the above-listed requirements can often be part of several customer communication-oriented applications such as eCommerce product listings, technical support knowledge-base, customer experience reviews, customer service, and more.

As the world inches closer towards digital technology as a part of daily life with each passing day, it is of utmost importance that businesses embrace technological tools to process and manage the most relevant content to accomplish its missions. eDiscovery is one such information triage application that, when combined with Machine Translation, brings amazing results. eDiscovery, when combined with MT, happens to be a crucial need that builds momentum as we become more digitally-focused workers.

But what really matters in an MT solution for eDiscovery?

In this post, we will be discussing the features of MT in eDiscovery that matter the most to active users based on the insights that we have gained from more than 50 years of powering translation for our eDiscovery clients.

### What is eDiscovery?

Electronic discovery (eDiscovery) is the process of identifying, collecting, and producing electronically stored information (ESI) as a response to some specific request for production in an internal corporate investigation or lawsuit. Some typical forms of ESI include emails, presentations, documents, databases, audio and video files, voicemail, website, and social media content.

The top advantages of eDiscovery include its dynamic nature. Unlike hard-copy evidence, eDiscovery evidence exists digitally which contains time-date stamps, file properties, and author and recipient information. By preserving the original content and metadata of electronically stored information, it further eliminates claims of spoliation or tampering.

With an ever-increasing digital world, greater and greater amounts of evidence exist in the digital format.

## The Power Couple: Machine Translation & eDiscovery

In an eDiscovery scenario, a combination of activities like classification, clustering, summarization, and N-Grams help in organizing and identifying the important material from huge databases. After organization, collation, and identification, it is likely the documents will need to be sent for translation. This where MT comes to help because of the sheer volume.

MT identifies the right document for refinement leveraging human translation. This process of identifying a small set of important documents from a large mass is basically the crux of the triage process.

Languages in eDiscovery are quite diverse and a lot of work goes into translating different source languages into English and sometimes German. Though people state that CJK and FIGS matter the most in this world, the needs vary from case to case – even Greek, Spanish, and Norwegian may happen to be important in certain cases.

Furthermore, when it comes to a particular business domain, patent infringement, litigation scenarios, and product liability dominate the lot. Other domains like consumer electronics, finance, IT, medical equipment, and automotive industries are also equally important.

## What Really Matters in an MT Solution for eDiscovery?

### Quick and Direct Accessibility

If there is one factor that attorneys, as well as corporate governance and compliance professionals value, the most when working with an eDiscovery platform, it is the ease with which they can operate MT. In most cases, they want to quickly and directly run document analysis and work on organization platforms. Though large documents can be fed into the MT in bulk, the ability to manage and review the important documents is again a crucial requirement.

### Language Identification

One of the first critical few steps in classifying documents includes organizing the documents based on their source language. Being the first level of triage, this step needs to be easy and efficient in order for the entire eDiscovery process to be smooth and hassle-free. Furthermore, some languages will also require different processing flows and non-automated procedures in case MT isn't available.

Reviewers are bound to follow only relevant threads and require ad-hoc translations of documents. Therefore, the MT should be capable of identifying the source language for a wide array of languages. Typically, reviewers will feed in a batch of documents in different languages and the MT solution should automatically identify and translate it.

### Processing Multiple Languages in One Document

From emails to office documents and social media to web content, all eDiscovery data is typically processed in a review platform such as Relativity. Often, an email thread can happen to exist in more than two languages. Thus, there arises a need for MT solutions in the market to handle multiple languages within the same document.

### **Security and Data Privacy**

Though users believe that systems installed on-premise with no data transported outside a secure firewall is safe enough, sometimes projects do come with data custody restrictions. This can limit the use of MT solutions dependent on the unique requirements of the user.

### **Ability to Process Large Data Sets Along with Ad-Hoc Needs**

Some projects include terabytes, even petabytes, of data that needs to be processed. In such cases, it is crucial to consider the raw-processing efficiency and performance of the MT solution. On the other hand, there might also be ad-hoc projects that contain data sets that are comparatively smaller in size. Therefore, MT solutions should provide a range of services that meet different user requirements. The degree of automation should be such that it can process 10,000 documents with the same ease as processing 10 documents.

### **Customization**

The complexity of customizing MT solutions vary based on requirements. For instance, customization is easily possible when it only includes general glossaries and dictionaries. However, rapid customization is a common case in eDiscovery where an MT application must use specific domain glossaries and focus engines relevant to the case. Integrations like these build a higher quality MT system that helps to extract the most relevant set of documents with minimum human translation efforts.

### **Integration with an eDiscovery Platform**

A robust MT solution must do more than have the ability to pass source code and process large files. It must have a native integration with a platform where eDiscovery is already happening. Relativity, an excellent document review platform that works closely with eDiscovery, is the preferred choice of many professionals to process multilingual content, especially in litigation scenarios.

### **Special Features**

Other than the features mentioned above, user-specific features — such as the ability to do anonymization, run corpus analysis and modification, and handle digital documents like audio and video files — can also be integrated into MT solutions; an important feature in today's "connected" world. For example, "smart" devices are increasingly impacting personal and business life. The data collected by these devices may be subpoenaed and included in evidence that needs to be processed during the discovery phase. MT applications can be configured to transcribe audio into text, translate the text into specific target languages, and remove sensitive data from the evidence so it can be analyzed while remaining in compliance with regulations like the GDPR.

### Meet SYSTRAN

Being a pioneer in the MT arena, [SYSTRAN](#) strives to address all the specific issues that really matter to an eDiscovery user. Like Relativity for eDiscovery applications, SYSTRAN is for multilingual eDiscovery. As a leading global MT solution provider in the eDiscovery segment, SYSTRAN showcases an excellent track record of success in the arena. Now, with SYSTRAN's partnership with Relativity, you get both solutions in a single platform.

### Introducing CMLess

The reason why SYSTRAN dominates the eDiscovery segment is that they partner with Relativity and own a native Relativity connector to bring the world [CMLess](#).

Built to feature the best practices of Relativity, SYSTRAN CMLess for Relativity provides a vital solution for customers in real-world multilingual discovery cases. The deep integration of the platform enables single- and multiple-language identification and translation within a single document. Consider the value if applied to a series of email threads.

**CMLess for Relativity**