SDL
Part of the RWS Group

Guide to Evaluating Machine Translation Software

Content Intelligence

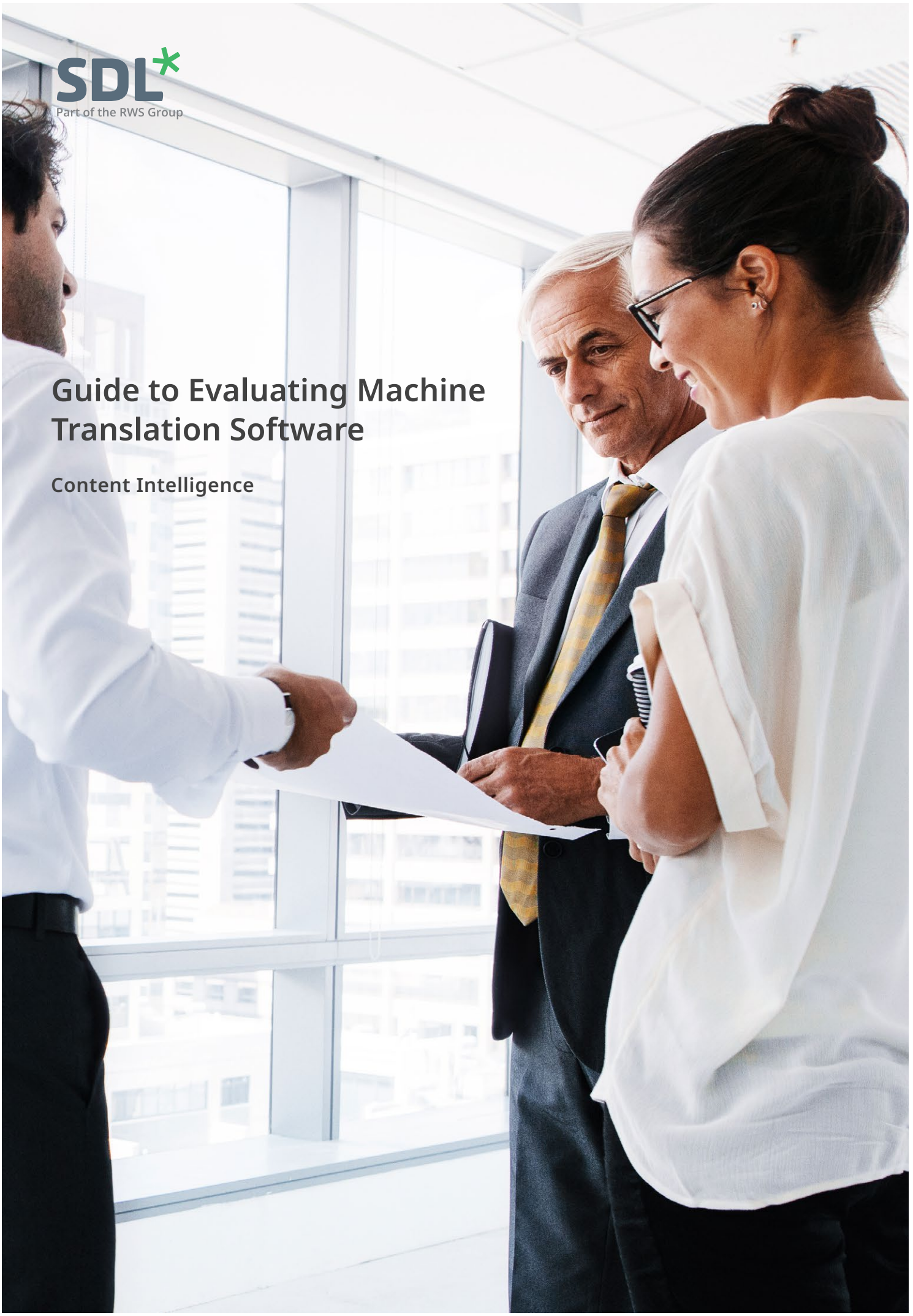


Table of contents

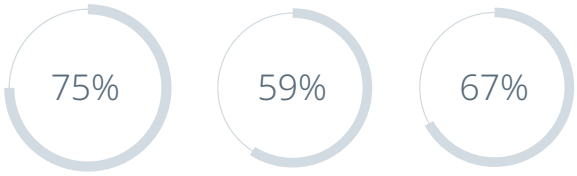
The Increasing Importance of Content	3
All Markets All the Time	4
The Limitations of Generic MT	5
Current Approaches to Evaluating MT Systems	7
A More Holistic Approach to Evaluation	10
Pervasive Translation Capabilities Drive Global Success	13
Key Attributes of a Superior MT Solution	14
The Unique SDL Offering	18
Looking Beyond MT	20
MT Technology Evaluation Checklist	22

The Increasing Importance of Content

Content is the infrastructure that connects civilization to commerce, innovation, and for the organizations who know how to harness it, dominance in the digital world. It not only drives the modern digital customer experience, but also is critical to business intelligence and inter- and intra-enterprise collaboration.

The issue of digital transformation is increasingly recognized as a primary concern and a key focus of executive management teams in global enterprises. The stakes are high for businesses that fail to embrace change. Since 2000, almost half (52%) of Fortune 500 companies have either gone bankrupt, been acquired or ceased to exist as a result of digital disruption. Once iconic brands now fall by the wayside and will quietly disappear, e.g., Sears, Toys "R" Us and many more. Studies suggest that many more companies across many industries will disappear because they fail to understand the strategic value of content and the related disruption it causes. Insight Research estimates that 75% of today's S&P 500 will be replaced by 2027.

Digital disruption supplies the consumer with better customer experiences by enabling the ability to deliver the right goods without ever having to leave one's house, wherever that house happens to be. The borderless, always on nature of the digital marketplace means all content must also always be global or else customers won't buy, as you can see from this research data from Shopify:



Want to buy products in their native language

Rarely or never buy from English-only sites

Prefer navigation and content in their language

Local language content makes or breaks global sales

The behavior of the modern customer has changed and is now highly impacted by freely flowing content. In fact, in many B2C and even B2B scenarios, the modern customer may conduct the whole customer journey without ever talking to a salesperson. Studies show as much as 67% of the buyer's journey and customer behavior is driven by content that they discover. As a result, companies that provide relevant, high-quality content succeed, and those that don't become irrelevant. The Impact of Digital Transformation graphic below shows the many stages where relevant customer content is required to persuade and connect with a potential customer and then maintain an ongoing relationship with an enterprise or a brand after a customer relationship has been established.

When companies can translate every asset, every video and every website into as many languages as technology will allow, companies will be present in every market because translation can be done quickly and affordably with the help of machines. Therefore, true globalization isn't just about generating revenue in another market. Being omnimarket, or the ability to translate all content for every language, is now the new "going global." Machine-first translations make content available in every language so arguing about whether translation is worth the investment to translate or not is no longer relevant.

Rather than localizing top-tier content exclusively for established markets, omnimarket organizations translate everything for everyone and see where the chips fall.

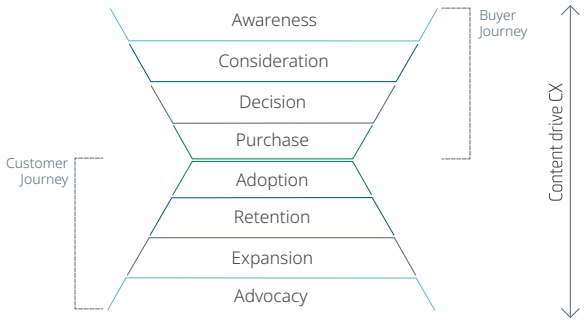
Content as a Driver of Global Digital Transformation

The Impact of Digital Transformation

Customers expect large volumes of relevant content to be available across all digital channels 24/7.

Content is the best salesperson for a digitally active and savvy customer.

Rapidly delivering the right content is key to being digitally relevant.

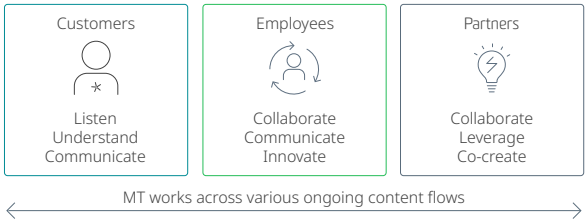


All Markets All the Time

Responding effectively to the realities of the borderless omnimarket world is now a matter of survival and a long-term competitive advantage.

In the modern era where customers interact continuously with enterprises on a variety of digital platforms, large volumes of relevant content are needed to support and enhance the customer journey. To remain relevant in this environment, enterprises must provide a continuous flow of content to all markets in all languages, regardless of volume. Whereas globalization only required a company to operate in a few targeted markets, an omnimarket strategy dictates that all markets be addressed in the borderless digital world.

MT Makes All Content Instantly Multilingual



What allows an omnimarket strategy to even be considered are technological advances in artificial intelligence. Machine Translation (MT) technologies enable organizations to make large volumes of content available to global customers, to understand what global customers are saying about brands in social forums and to have multilingual conversations with their customers. As a result, machine translation has evolved business-related translation requirements beyond the traditional focus of localization for just a few key pieces and languages. The modern enterprise need for translation involves many millions of words a day to expedite communication and collaboration within the global enterprise and with international business partners. Sharing content is no longer just for improving customer experience; it is now critical for improved global collaboration and better business intelligence.

It is estimated that multiple trillions of words are being translated across hundreds of language combinations every day using generic, public machine translation. While generic web users drive much of this volume, it is now understood that a significant portion of this astounding volume is from enterprise employees who use public MT to translate confidential enterprise content to respond to urgent global market communication and collaboration needs.

Demand has already reached a point where it is not viable to service this need without using technology. However, public MT technology creates significant risk in terms of data security and privacy and introduces inefficiencies into the modern global enterprise.

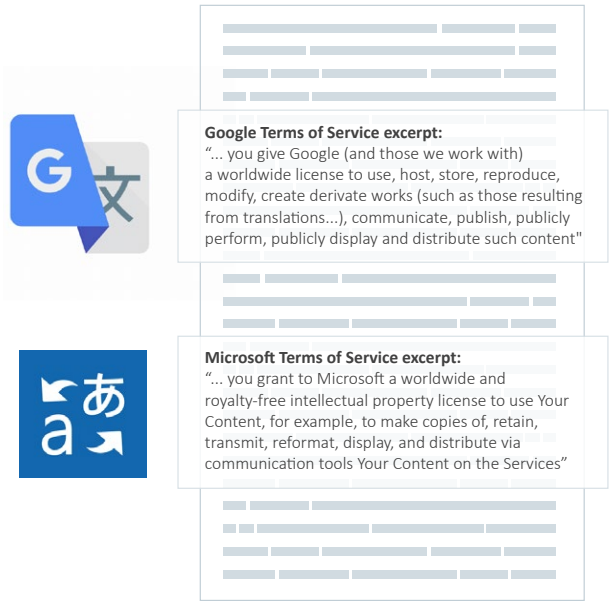
The Limitations of Generic MT

Hundreds of millions of people across the global internet regularly use machine translation on a variety of public MT portals. MT is no longer a fringe case, but now a mainstream utility.

However, the use of public MT presents some serious hurdles for the enterprise, especially those that want to embed MT capabilities into their core global communication and collaboration workflows or use MT to understand and analyze sensitive customer feedback data. Public MT portals also present a potentially significant data security risk for an enterprise.

"The fundamental truth highlighted by Schneier's aphorism is that the vast majority of internet users have entered into a Faustian bargain in which they exchange control of their personal data in return for 'free' services (such as social networking, MT and search) and/or easy access to the websites of online publications, YouTube and the like."[1]

Data Security



[1] theguardian.com/commentisfree/2018/apr/08/dont-blame-facebook-taking-data-most-pubishers
[2] policies.google.com/terms?hl=en
[3] microsoft.com/en-us/servicesagreement/default.aspx
[4] conversationagent.com/2018/04/05/index.html

The common understanding is that the MT services are "free", but as with many "free services" on the internet today, the providers of the free services monetize their technology by harvesting user data. This data can then be used to target users with specific advertising messaging and generate more revenue.

The Google Terms of Service very clearly states the following:

"When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones."

We regularly hear about web services experiencing data breaches and inadvertent data leakages. The **Facebook Cambridge Analytica** saga and recent fines for **Google for using data to target ads to children** are only the most recent and visible incidents. The practice of data harvesting is extensive and pervasive.

As Valeria Maltoni stated:[4]

"Breaches expose information the other way. **They shine a light on the depth and breadth of data gathering practices – and on the business models that rely on them**. Awareness changes the perception of knowledge and its use. Anyone not living under a rock now is aware that we likely don't know all the technical implications, but we know enough to start making different decisions on how we browse and communicate online.

Business models are the most problematic, because they create dependency on data and an incentive to collect as much as possible. Beyond advertising, lack of transparency on third party sharing and usage merit further scrutiny. Perhaps the time has come to evolve business practices – how platforms and people interact – and standards – based on laws and regulations."



A particularly egregious example of what can go wrong is demonstrated by the data exposed when employees of Norwegian state-run oil giant Statoil "discovered text that had been typed in on [translate.com] could be found by anyone conducting a [Google] search." This text included **"notices of dismissal, plans of workforce reductions and outsourcing, passwords, code information and contracts".**

There have been recent reports that MT services hosted in foreign countries and utilized by international users may be monitored by local government agencies as part of standard surveillance activities. As public MT services proliferate around the world, it is essential for all enterprise users to take special precautions when utilizing any open MT resource where this risk exists. Government sponsored data collection and surveillance are likely to gain further momentum in the future as cybersecurity becomes increasingly equated with national security.

The fact of the matter is that today the possibility of highly confidential and privileged information being passed through an MT portal is very high if special care is not taken to prevent this. Data security has to be specifically planned for and implemented using technology alternatives that are designed from the outset to provide and ensure this.

Adaptation – Optimization for the Enterprise

Public generic MT is not well suited for enterprise purposes because it is a one-size-fits-all approach to MT. A single MT engine will service any and all translation requests in that specific language combination. Most MT experts will tell you that MT performs best when it is optimized, tuned for a specific linguistic profile and trained with relevant terminology.

Ensuring the MT technology performs well on the primary content and focuses on the enterprise's products and business affairs is a critical requirement for successful enterprise MT deployment.

MT system capabilities may also vary depending on the business purpose. For example, a system optimized for internal employee communications may not be appropriate to translate a technical knowledge base.

5 aspi.org.au/report/engineering-global-consent-chinese-communist-partys-data-driven-power-expansion

Thus, enterprises need MT technology that can be updated and tuned for their specific business needs, typical terminology and business language.

Public Generic MT Systems vs. Adapted Enterprise MT Systems

Public Generic MT Systems

- Goal is to get a general understanding
- Basic quality translation intended for wide applicability
- Focus = Broad but shallow
- Google, Bing, Baidu, DeepL and other free sites
- One size fits all
- Potential loss of ownership, data security and privacy
- Limits to volume translated

Adapted Enterprise MT Systems

- Usually built with customer data
- Much higher accuracy and translation quality
- Focus = Narrow but deep
- Optimized for a specific customer-defined domain
- Matched to a specific business purpose
- Secure data, private system
- No volume limits
- Complete control and integration into enterprise IT

Operational Efficiency

Public generic MT solutions often require users to cut and paste content they want translated. When this is multiplied across several instances and many thousands of workers, it is clearly a sub-optimal way to use MT technology.

When closely integrated with the global enterprise's content management infrastructure, enterprise language optimized MT is a much more efficient and effective way to implement MT. If done securely, it enables global communication and collaboration that allows the enterprise to participate in digital transformation.

Current Approaches to Evaluating MT Systems

As the value of freely flowing and relevant content continues to rise, it becomes increasingly clear that MT is a wise and needed technology. If properly deployed, MT can ensure that the enterprise is responsive and actively able to participate in a global marketplace with reduced language-based friction.

Technology buyers are often confused about how to select the "best" MT technology for their use. They often focus on dubious proxies for "translation quality" rather than looking at the overall suitability and infrastructure to mold MT systems to unique enterprise needs. For instance, in the selection process many buyers consider MT technology rankings based on a single "quality score" without understanding what these scores really mean and what they actually measure. The most common way to measure quality is to compare the output strings of automated translation to a human translation of the same sentence. The issue with this score-based approach is that these scores are not true quality indicators.

Measuring translation quality is difficult because there is no absolute way to measure how "correct" a translation is. MT is a particularly difficult AI challenge because computers prefer binary outcomes and translation rarely - if ever - has only one single correct outcome. 2+2=4 but so does 3+1 and both are correct.

The fact that one human translator will translate a sentence in a significantly different way than another human translator leads to problems when using these human references to measure "the quality" of an automated translation solution. In the translation world, there are as many "correct" answers as there are translators.

BLEU Score

BLEU (Bilingual Evaluation Understudy) is the most common output quality measurement used in the MT research community. Many of the newer derivatives of BLEU (METEOR, LEPOR) are very similar and attempt to measure precision and recall in an objective way. However, they only measure the text-string similarity of MT output to one or more human translations of the same material.

The primary use of automated metrics like BLEU is to provide feedback to MT system developers during the system development process as they modify data and development strategies. In that context, these metrics can be very useful.

However, when used to compare MT systems, these metrics become much less useful and accurate.

Human	Automated Metrics
• Segment Rating • Error Classification • Edit Distance	• BLEU • F-Measure • METEOR • Rouge • LEPOR • ChrF
Slow, costly and inconsistent	Rough approximations, not true quality indicators, string matching, need proper protocol

Problems in Comparing MT Systems with BLEU and LEPOR

The BLEU score is a string-matching algorithm that provides basic output quality metrics for MT researchers and developers. A fundamental problem with BLEU is that it **does not even try** to measure "translation quality", but rather focuses on **string similarity,** usually to a single human reference. Over the years people have chosen to interpret this as a measure of the overall quality of an MT system. BLEU scores only reflect how a system performs on the specific set of test sentences used in the test. As there can be many correct translations and most BLEU tests rely on test sets with only one correct translation reference, perfectly good MT translations can be scored poorly.

BLEU is actually nothing more than a method to measure the similarity between two text strings. To infer that this metric, which has no linguistic consideration or intelligence whatsoever, can predict not only past "translation quality" performance but also future performance, is indeed quite a stretch.

It is critical to understand that **the scores do not reflect the potential performance of the system on other material** that differs from the specific test material. All inferences on what the score means should be made with great care **after taking a close look at the existing set of test sentences.** It is very easy to use and interpret BLEU incorrectly and many examples of incompetent use can be found throughout the translation industry.

Also, it is important to run these tests using material that is clearly understood and not previously seen by the MT system. If the MT system has seen the test data before and it was used to build the system, the test results and score will be meaningless. It would be the equivalent of testing a student with an exam where the answers are also provided. Many of the public MT system rankings we see today are unable to confirm that the test data has not been used in training, making the rankings quite suspect.

Multiple Translations Can Be Technically Correct. Ten Translations of a Chinese Sentence

A typical example from the 2001 National Institute of Standards and Technology (NIST) evaluation set:

这个 机场 的 安全 工作 由 以色列 方面 负责。

Israeli officials are responsible for airport security.

Israel is in charge of the security at this airport.

The security work for this airport is the responsibility of the Israel government.

Israeli side was in charge of the security of this airport.

Israel is reponsible for the airport's security.

Israel is responsible for safety work at this airport.

Israel presides over the security of the airport.

Israel took charge of the airport security.

The safety of this airport is taken charge of by Israel.

This airport's security is the responsibility of the Israeli security officials.

BLEU in Action - A French Sentence Translation

Source Original:
Le professeur est arrivé en retard à cause de la circulation.

Reference Translation:
The teacher arrived late because of the traffic.

Many accurate and correct translations can score lower simply because they use different words.

Automated quality measurement metrics have always been important to the developers and researchers of data-driven based MT technology because of the iterative nature of MT system development and the need for frequent assessments during system development. These metrics can provide rapid feedback on the effectiveness of continuously evolving research and development strategies.

The professor was delayed due to the congestion.	#1 Very low BLEU score
Congestion was responsible for the teacher being late	#2 Slightly higher but low BLEU
The teacher was late due to the traffic.	#3 Higher BLEU than #1 and #2
The professor arrived late because of circulation.	#4 Higher BLEU than #3
The teacher arrived late because of the traffic.	#5 Best BLEU Score

Recently, we have seen that BLEU and some of its close derivatives (METEOR, NIST, LEPOR and F-Measure) are often used to compare and rank the quality of different MT systems in enterprise use settings. This can be problematic since a "single point quality score" based on publicy sourced news domain sentences is not representative of the dynamically changing, customized and modified potential of an active and evolving enterprise MT system. **Also, this type of score does not consider the importance of overall business requirements in an enterprise use scenario where other workflow, integration and process-related factors may be more important than small differences in scores. Useful MT quality in an enterprise use context will vary greatly depending on the needs of the specific use case.**

Common Problems with BLEU-Based Comparisons

- **Wrong Domain:** Cross-vendor comparisons are most often based on publicly available News Domain data (WMT data)
- **Not Blind:** The Test Set has been seen and used in training by some of the MT systems being compared
- **Size:** The Test Set is too small
- **Inadvertent Bias:** Many researchers have noted that Test Sets using public MT and post-edited introduces a preference bias for string matches
- **Static:** MT systems evolve continuously in the enterprise as they are adapted to specific use cases

While BLEU is very useful to those who build and refine MT systems, its value as an effective way to compare totally different MT systems is much more limited and needs to be used very carefully, if at all, as it is easily and often manipulated to create the illusion of false superiority.

"CSA Research and leading MT experts have pointed out for over a decade that these metrics [BLEU] are artificial and irrelevant for production environments. One of the biggest reasons is that the scores are relative to particular references. Changes that improve performance against one human translation might degrade it with respect to another. Approaches that emphasize usability and user acceptance take more effort than automatic scores, but point the way toward a useful and practical discussion of MT quality."
Excerpt from a CSA Blog on BLEU Misuse, April 2017

In competitive comparisons, it is important to conduct the tests in an unbiased, scientific manner to get a true view of different alternatives. The "test set" should be unknown ("blind") to all the systems that are involved in the measurement. This is often violated in many widely used comparisons today. If a system is trained with the sentences in the "test set", it will obviously do well on the test but probably not as well on data that it has not seen before. Many recent comparisons score MT systems on news domain-related test sets that are easily obtained from the research community but may also be used in training by some MT developers. A good score on news domain may not be especially useful for an enterprise use case that is heavily focused on IT, pharma, travel or any domain other than news.

One problem with using news domain content is that it tends to lack tone and emotion. News stories discuss terrorism and new commercial ventures in almost exactly the same tone. In business communication, customer service and support, tone really matters. Enterprises that can identify dissatisfied customers and address their concerns are likely to be more successful. Improved CX is all about understanding tone and emotion in addition to the basic literal translation.

In spite of the limitations identified above, BLEU continues to be a basic metric used by most, if not all, MT researchers today. However, most expert developers regularly use human evaluation on smaller sets of data to ensure that they indeed have a true and meaningful BLEU. The MT community has found that purportedly improved metrics like METEOR and LEPOR have not really gained any momentum. BLEU and its flaws are more clearly understood, and thus more reliable, especially if used together with supporting human assessments.

Many buyers today realize that MT system performance on their specific subject domains and translatable content for different use cases matters much more than how generic systems might perform on news stories.

Also, many sophisticated users understand that small differences (+/- 15%) in these scores do not make one system better or worse. What matters most is the business impact and outcomes of properly deployed MT solutions, which very rarely depend on higher BLEU or LEPOR scores of the base MT system.

People want "just one measure", or a single letter grade to understand system quality, but the actual reality is much more complex than that.

The danger of single measures that summarize complex performance characteristics is that they are almost ALWAYS inadequate.

It should also be understood that most MT systems used in enterprise settings can be continuously updated and modified to improve system performance on the changing language and terminology that is most commonplace in enterprise content. New terminology is added as products and customer situations evolve. A single point score in many senses is just a photo of a dynamically moving and changing scene, and therefore of limited value.

A More Holistic Approach to Evaluation

The issues discussed challenge the notion that single-point scores can really tell you enough about long-term MT quality performance. This is especially true as we move away from the localization use case. Speed, overall agility, responsiveness and integration into customer experience-related data matter much more in many business applications of MT. The actual translation quality variance measured by BLEU and LEPOR may have little to no impact in the following business applications, where MT enhances and accelerates global content and communication flows to improve customer experience.

Enterprise MT Use Cases


Global Communication and Collaboration


Global Customer Care and Support


Understanding the Global Customer

The enterprise value equation is much more complex than linguistic quality and often improperly calculated Natural Language Processing (NLP) scores. The key enterprise requirements for deploying MT technology are overall control and adaptability. To truly reflect the business value and impact, evaluation of MT technology must consider non-linguistic attributes including:

- Adaptability to business use cases
- Data security and privacy
- Integration into enterprise infrastructure
- Deployment flexibility and control
- Clearly identified business outcomes that may have little or nothing to do with the linguistic quality of the output

To effectively link MT output to business value, we need to understand that although linguistic precision is an important factor, it often has a lower priority in high value business use cases. This view will hopefully take hold as the purpose and use of MT is better understood in the context of larger business impact scenarios beyond localization.

A More Meaningful Evaluation Framework

While single-point scores do provide a quick and dirty sense of an MT system's linguistic quality performance, it is more useful to focus testing efforts on specific enterprise use case requirements. This is also true for automated metrics, which means that scores based on news domain tests should be viewed with care since they are not likely to be representative of performance on specialized enterprise content.

The most important factors to consider when rating different MT systems are:

- **Adaptability options:** Range of options and controls available to tune MT system performance for very specific use cases. For example, optimization techniques applied to eCommerce catalog content should be very different from those applied to technical support chatbot content or multilingual corporate email systems.
- **Data privacy and security:** If an MT system will be used to translate confidential emails and business strategy documents that are only consumed by internal stakeholders, data security and privacy will be more urgent than for localization content. Buyers should understand that many public MT systems will harvest all user data for machine learning purposes to "improve the service". When MT is implemented to enhance global communication and collaboration, public systems introducing potentially real risks can be avoided.
- **Deployment flexibility:** Some MT systems need to be deployed on-premises to meet legal requirements, such as in litigation scenarios or when handling high-security data. The availability of both cloud and on-premises options, or combinations thereof, provide flexibility to an enterprise with varied deployment needs.
- **Expert services:** Having highly-qualified experts to assist with MT system tuning and customization can be critical for certain customers to develop optimal systems and ensure successful business outcomes.
- **IT integration:** Increasingly, MT systems are embedded in larger business content workflows to enable instant multilingual capabilities, for example, in communication and collaboration software like email, chat and CMS systems.
- **Overall flexibility:** Together, all these elements provide flexibility to tune the MT technology to specific use cases and develop successful solutions.

Business Impact

The integrity of the overall solution likely has much more impact than the MT output quality in the traditional sense. Not surprisingly, MT output quality could vary by as much as 10-20% on either side of the current BLEU score without impacting the actual business outcomes. Linguistic quality matters, but is not the ultimate driver of successful business results. In fact, improvements in output quality in an eCommerce use case actually reduced the conversion rates on post-edited sections, as this post-edited content was viewed as being potentially advertising-driven and thus less authentic and trustworthy.

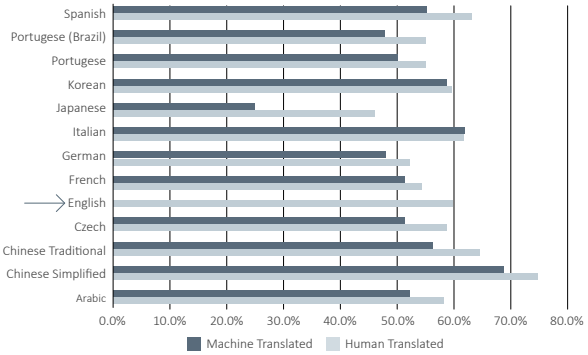
In the context of MT for technical support, the primary concern is whether the content is easily found, relevant and can help customers solve technical support problems:

- Is MT "good enough" for the **live agent** to understand?
- Is MT "good enough" for **the customer** to understand?
- Is the **Customer Experience** improved by this?

In the case referenced below, a common linguistic quality assessment process is used across all languages for knowledge base content. In most cases, extensive MT is used to make all the source English content multilingual. The executives managing technical support determined business success by measuring responses to two questions: "was this information relevant?" and "did this information solve your problem?" While attempts are made to reach the highest acceptance levels in terms of translation quality, the reality is that often executives have to choose between providing support content using MT or nothing at all. Evidence suggests that most users prefer MT content over trying to decipher the source language. The following chart shows MT-translated content produces almost the same success rates for self-service technical support as human translation.

Customer Self-Service: Success Rate by Language

"Success rate" combines two sets of numbers: **"was this information relevant?" and "did this information solve your problem?"**



Source: Martine Smets, Microsoft Customer Support @GALA 2016

This chart also shows that despite the same linguistic quality assessment rules being applied to all the languages, the results may differ for both human and MT content.

Thus, more useful measurements for MT's impact and value might be as shown below:

MT Quality for Customer Service and Support

- Higher volume of successful self-service
- Easy and rapid access to all new support content
- Increased global customer satisfaction
- Expanded global digital presence

If we consider MT's use in enabling and improving global communication and collaboration both inside and outside the enterprise, we can see that there are other measures for success.

Assuming there will be substantial volumes of constantly changing communications content, the possibility of getting perfect human-equivalent translations is nil given the volume and velocity of the flowing information. In this use context, MT enables organizational agility and responsiveness. Also, in this setting, data security and seamless integration with core IT infrastructure is perhaps even more critical than linguistic quality.

Indicators of Successful Business Outcomes for Different MT Use Cases

It is important for enterprise technology buyers to understand that successful outcomes with the use of MT technology are not strictly tied to the linguistic output quality produced by an MT system. While achieving the best possible output quality is an important goal, other factors like the quality of the integration, content findability, relevance and customer user experience-related issues may play a much bigger part in producing successful outcomes. With high-volume volatile information translation done by MT, all MT is likely to fall short of competent human translation. However, if the translation is useful to the business purpose despite being linguistically flawed, it has served its business purpose. The variance in output quality can be as high as 10%, or even 20%, and still produce the same business outcome. The fact is that the BLEU score of an MT system has a very limited impact on the business outcome in the dynamic content flows that characterize modern digital communication and collaboration. It is possible that MT systems with lower scoring output quality as measured by a BLEU score can outperform higher scoring systems if they are properly integrated and implemented.

Technology implementation needs to consider business outcomes above all else as described below in various MT use case scenarios.

Global enterprise communication and collaboration

- Increased volume in cross-language internal communication and knowledge sharing with safeguarded security and privacy
- Better monitoring and understanding of global customers
- Rapid resolution of global customer problems, measured by volume and degree of engagement
- More active customer and partner communications and information sharing

Customer service and support

- Higher volume of successful self-service across the globe
- Easy and quick access to multilingual support content
- Increased customer satisfaction across the globe
- The ability of monolingual live agents to service global customers regardless of the customer's language

eCommerce

- Measurably increased traffic drawn by new language content
- Successful conversions in all markets even with imperfect MT content
- Transactions are driven by newly translated content
- The stickiness of new visitors in new language geographies
- Social media analysis
- Ability to identify key brand impressions
- Easy identification of key brand themes and issues
- A clear understanding of key positive and negative reactions

Use Case Optimization

MT quality needs to be tied to overall utility and business purpose. In general, the higher the volumes the more difficult it is to achieve human translation quality. However, in many AI initiatives, including MT, it is not necessary to get 100% accuracy for the technology to deliver value.

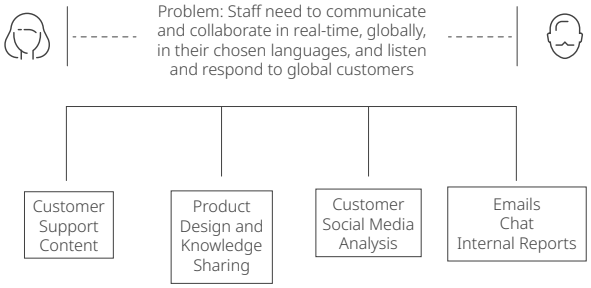
MT Quality Assessment to Match Business Purpose

Content Type	Target Quality	Process	Volumes
Legal, Marketing, Localization	High	Extensive PEMT	Low
Technical Support, Knowledge Base	Moderate	Custom MT + Linguistic Steering	Moderately High
Customer Experience Content	Moderate to Low	Custom MT + Linguistic Steering	Very High
Corporate Communication and Collaboration Content	Moderate	Custom Corporate MT	High

Pervasive Translation Capabilities Drive Global Success

For a global enterprise seeking to rapidly increase communication with its global customers, partners and employees in internal departments like product, engineering, support and marketing, instant translation is increasingly a key enabling tool. However, MT solutions need to be powerful, secure and flexible to adapt to changing content type, format, linguistic style and target deployment requirements. The modern enterprise needs to be continually expanding its dialogue, monitoring and communication with its customer base as broadly as possible.

Where Can Translation Be Used in the Enterprise?



In a typical global enterprise, the translation need goes beyond making more content available to global customers in many languages. There is an ongoing need for collaborative information sharing between multilingual global teams and increasingly for daily communications about customers, content enhancement and rapidly evolving business intelligence.

Content Drives Revenue and is Critical to Overall Customer Experience



The best MT technology will be able to adapt and support these very different use scenarios while enabling rapid reconfiguration of MT technology capabilities to support new use cases that may emerge to improve international market-related activities. The most pervasive and actively used MT technology today (Phrase-based Statistical Machine Translation (SMT), Adaptive MT and Neural MT (NMT)) are all "data-driven" technologies, which means computers learn from historic human translation and other available linguistic data. **SDL is a pioneer in data-driven MT, holding many patents and, most importantly, 15 years of in-depth experience developing enterprise production-ready MT systems**. Key members of the original Google MT team were initially trained at SDL/Language Weaver. SDL is one of the largest enterprise-focused developers of private and secure MT solutions and, as of 2018, customized SDL cloud systems were translating 30 billion words a month. Additionally, on-premises systems provided by SDL are translating multiples of this volume.



Key Attributes of a Superior MT Solution

Data Security and Privacy

For a modern global enterprise, government agency or high-value content publisher, data security and privacy are critical concerns. Today, MT technology is pervasive and easily available in search portals and some social media, but the "free" use of these services also means that data security and privacy are compromised. Public MT portals have sophisticated data harvesting infrastructure in place and analyze and use all user data to "improve the service."

Translation data related to important customer and partner interactions in the sales and support framework or internal communications about new product and service initiatives are generally considered highly confidential information. Few enterprises would want their confidential information accessible to machine learning algorithms that continuously crawl data in the open internet space. Important data security and privacy needs of the global enterprise can be addressed in two ways:

1. Utilizing "on-premises" translation servers, which may also include private, certified clouds or any customer-defined computing host environments that have a secure and authorized-use-only ability to deploy specialized and private MT engines. Control of the computing environment to run translations is a key requirement for many enterprises. As dialogue with customers increases, this control will become increasingly more important.
2. Contractual protection on the appropriate use and re-use of any and all private data provided to a MT vendor to train and build specialized customized MT systems for an enterprise customer.

Typically, public generic use systems are not designed for the very specific domain needs and integration requirements of a global enterprise and allow very limited, if any, adaptation of the generic engines to the continuously changing needs of a global enterprise.

Technical Competence and Expertise with MT Technology

Machine translation is a complex technology that is constantly evolving. Successful use requires collaboration with a vendor that has all of the following:

- Deep understanding of how MT works to adjust, modify and adapt generic functionality for different use scenarios.
- Long-term experience in many real-life use scenarios, support tools and utilities to tune systems for specific and unique requirements.
- A technical team knowledgeable in the latest research and breakthroughs, with backgrounds in computational linguistics, corpus analytics and preparation and machine learning. Regular collaboration between the MT technology team and linguists is particularly advantageous.
- Competence with computing system architecture and IT requirements that have to scale and change according to different use scenarios and enterprise needs.
- A comprehensive suite of NLP and data transformation tools to allow rapid configuration modifications for different use scenarios.

A combination of all of the above factors will result in MT systems that can be quickly and easily adapted to produce the highest quality translation output.



High-Quality Data

Given that all emerging machine translation technology models today are trained by linguistic data sets, it is critical to have high-quality data as a foundation for accurate learning to take place. Data-driven technologies are particularly affected by the GIGO (Garbage In, Garbage Out) principle. If MT engines are trained on noisy, dirty data, translation quality will be compromised and the value of the MT output will be greatly reduced. The following factors also matter:

- Long-term experience with data-driven methodology and regular contact with linguists ensures higher-quality data resources and greater ongoing leverage.
- Availability of specific domain data sets (vertical industry e.g., IT, Automotive, Legal, Business Correspondence, Travel) to accelerate custom engine development
  or higher quality domain output prior to customer data customization.
- Extensive metadata on all data assets to enable accurate and granular extraction of pertinent data for different domain adaptation and use case needs.
- Availability of corpus analysis and cleaning tools ensuring all format, linguistic characteristics and critical terms are correct and consistent in the training data.
- Availability of corpus comparison tools to identify data gaps, drive vocabulary expansion in a targeted way and build relevant data assets optimally.

Linguistic Steering

While MT technology continues to evolve and improve, the best MT output quality is typically achieved by close collaboration with linguists. Steering the MT development process, linguists provide valuable error feedback at a pattern level and ensure the data used to build the MT engine is consistent and clean. Linguists work with the MT development team to ensure that the output reaches the optimal level of adequacy and fluency for each use scenario. Linguists also assess MT output quality as an engine evolves to evaluate whether the development strategy needs adjustment since automated quality measures can sometimes be misleading. Some of the systematic tasks that linguists may perform include:

- QA on the translations of high frequency patterns in a source corpus to ensure consistent and accurate learning.
- Rationalization and normalization of critical terminology.
- Assessing MT for PEMT suitability and understanding productivity implications of different MT engines.
- Language-specific data management issues (handling of missing diacritics, morphology, widespread use of diminutives, etc.).
- Matching source and training corpus overlap and identifying and resolving vocabulary gaps to ensure better MT output quality.
- Training post-editors to maximize efficiency.
- Development of special runtime scripts to handle unique linguistic issues.
- Providing error-pattern feedback to developers to drive error correction strategies at a pattern-level rather than a segment level.

Language Coverage

While broad language coverage is always good, in global enterprise use cases language pairs are defined by very unique and specific needs. Some vendors claim large numbers of language combinations by merely calculating permutations with large sets of public data, such as the EU Europarl data. Systems built from these data either tend to be low value for enterprise needs or include obscure combinations like Danish to Latvian, which has little value in commercial enterprise deployment scenarios.

Domain-adapted generic systems, or vertical industry engines, can be very useful to accelerate the development of customer-unique engines or provide immediately useful tuned engines. Combinations not generally available can easily be created for a defined customer need. For example, a German auto manufacturer who also builds cars in Poland, Brazil and China may need to make German content rapidly available in each of these countries. Sufficient volumes of customer data can easily enable this. The most agile MT developers will have multiple ways to approach this challenge, depending on training data availability and data creation strategies using a variety of NLP tools and specialized human translation efforts.

In the enterprise context the number of total languages is less important than depth (domains) and quality of the base MT systems. Also, it is important to have a language that has commercial relevance, such as Hindi rather than Pashto, as many more commercial opportunities will be driven by Hindi, though Pashto may be very valuable in a Governmental National Security environment.

Deployment Options

Global enterprises have varying needs in terms of how they might deploy an MT solution. A vendor who provides multiple options would generally be preferred to one who can only provide a single, vendor-preferred option.

The available options are typically:

Secure Cloud

This is often the best approach for a customer who does not want the burden of managing their own computing resources. This option can sometimes mean that the vendor has access to client data and could use this data to "improve their services" as is stated openly by some public MT portals. The cloud is increasingly a preferred mode of deploying desired computing functionality as the management and protection of servers is a non-trivial task and often beyond the expertise of many SMEs and LSPs. Different vendors offer different kinds of performance and security guarantees but many run software services in Amazon and Microsoft's cloud offerings, which have very clearly-defined availability and security parameters. Some vendors have private and relatively secure cloud offerings.

On-premises

The most literal meaning of "on-premises" is a physical server in a client-controlled physical environment. Given the complexity of managing production servers, many enterprises may now have a "private cloud" that provides the same kinds of data security and privacy options as an actual on-premises server but has expert IT personnel managing the availability, protection and administration of the physical devices to meet today's 24/7 availability requirements.

It is also possible to set up private areas (non-shared server resources) in public cloud offerings like Amazon and Microsoft.

Hybrid Cloud

Some vendors can even combine these two deployment scenarios and enable the enterprise to set up some on-premises servers for high-use languages and make other less-used languages available in the cloud.



Integration with Organizational Source Data Flows

Enterprise organizations rarely use MT solutions only for the translation of simple phrases. While the average internet user may translate words and phrases occasionally, most enterprise and professional use of MT is embedded within a particular business information flow or software application, such as communication and collaboration software, Translation Management Systems (TMS) and TM systems or linked to web content like FAQs, knowledge bases and chat platforms for customer service and technical support.

This requires the MT solution to have a standard API so it can be easily connected to the required business software application. A good API will allow more than just source and target text exchange and may support other special requirements that are needed and useful for different applications.

The most important integration in a professional translation organizational setting is with TMS and/or TM software. MT is called upon when chunks of text need to be translated and typically the output will be post-edited.

When MT is used in a broader global enterprise setting to expedite cross-language communications and multilingual information sharing, integration with office productivity software like Office, email and other content sharing software is mandatory.

Many enterprises manage comprehensive product-related content in Content Management Systems (CMS). Often large volumes of information from these systems are sent to translation processes and then to a target website for customer access. The ability to link MT into these critical data flows is important for global enterprise as it makes implementing ongoing updates and maintenance more efficient.

Increasingly, a large amount of content passing through enterprise MT servers related to UX and product feedback is acquired from customer, partner and subsidiary comments on the website, which can enhance international business initiatives. The sheer volume and velocity of this type of information makes traditional human translation processes prohibitive.

For enterprises wanting to systematically monitor what active and vocal customers and influencers are saying about their user experiences on review and commenting platforms, MT systems may also require integration to expedite rapid ingestion and analysis.

Specialized applications like e-Discovery will require integration with the platforms where discovery documents are organized using metadata and analyzed and structured for use in specific scenarios like corporate litigation.

Professional Services

MT solutions providers who offer expert engine development and linguistic steering consulting are able to handle complex use scenarios and produce higher quality with more robust implementations. There are three areas where professional services can add value, specifically:

1. Engine development and use case optimization
2. Linguistic steering and linguistic data optimization
    a. Corpus management and training data cleanup and preparation
    b. MT quality assessment
    c. Post-Edit Machine Translation suitability assessment, training and development
    d. Error pattern communication with developers
    e. Rapid error correction strategies
    f. Runtime rules development to dynamically address error patterns and source/target issues without re-training engines
3. Integration with organizational workflows and software

SDL Enterprise MT Provides Control and Flexibility

Integrations	• Analytics • Office productivity • Collaboration • TMS • ASR/OCR
Control Depolyment and Data	• Data security • Data privacy • Confidentiality
Control Quality	• Flexible adaptability tools and services to drive quality to desired levels for each use case

The Unique SDL Offering

SDL is unique in the MT solutions marketplace for addressing the criteria listed above in the most complete and comprehensive manner.

Key Requirements of Enterprise MT

Low Cost On-premise, Rapid Adaptation, Deployment Flexibility

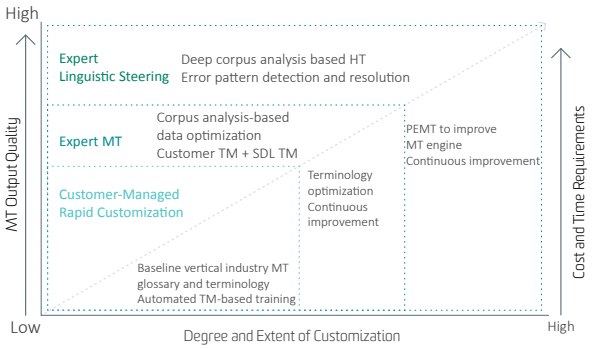
The SDL MT Focus: Key Enterprise Themes

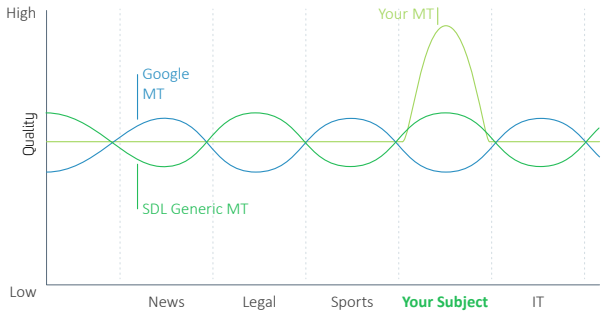
Quality Improvements	Lower TCO	Depolyment Flexibility	Straightforward Adaptability
Neural MT 2.0	NMT on CPU Hardware Based Optimization	On-Premise, Private Secure Cloud	NMT Dictionaries NMT Trainer

SDL today provides the most complete combination of the following:

- State-of-the-art MT technology initiatives including NMT, SMT and Adaptive MT
- Highest levels of data security and data privacy with both SDL Cloud and on-premise options
- Comprehensive and experienced professional services for both linguistic and MT engine development tasks
- High-quality underlying data to support and drive all MT solutions development
- Multiple deployment options – SDL Cloud, Private Customer Cloud, Public Cloud and On-premise - or any combination of these
- Broad range of enterprise-oriented capabilities to facilitate large-scale deployment

Tailoring MT Performance to Enterprise Needs





MT Customization and Tuning Impact on Quality

- SDL's Ongoing Focus on Quality Improvements
- Improving NMT dictionary features
- Tackling specific NMT challenges
- Adding new languages and new combinations
- Optimizing quality and speed
- Several do-it-yourself and expert-built MT systems options
- Broad range of integrations and API connectors for office productivity, CMS, TMS and TM systems

The SDL eco-system of translation, content management and language technology services enables SDL to deliver the most complete and most deeply integrated MT services in the industry.



Lower TCO
NMT on CPU



Docker Support

Collaborate
Communicate
Innovate

SDL MT Edge

Hardware-Based Optimization

Allows you to optimize the performance of your Neural MT Models (speed/quality) based on your hardware (CPU or GPU)

Neural machine translation has quickly established itself as the preferred model for most MT use cases today. Most experts now realize that MT performs best in industrial deployment scenarios when it is adapted and customized to specific subject domain, terminology and use case requirements. Generic MT is often not enough to meet key business objectives. However, successful development of adapted NMT models is difficult for the following reasons:

1. The sheer volume of training data required to build robust systems, typically in the hundreds of millions of words range that few enterprises will ever be able to accumulate and maintain. Models built with inadequate foundational data are sure to perform poorly and fail to meet business objectives or provide business value. Many AI initiatives fail or under perform because of data insufficiency.
2. Available options to train NMT systems are complex. Almost all of them require that any training data used to adapt NMT systems be made universally available to the development platform to further enhance their platform. This raises serious data security and privacy issues in today's digital era, where the data related to the most confidential customer interactions and product development initiatives need to be translated daily. Customer interactions, sentiment, service and support data are too valuable to be shared with open source AI platforms.
3. The cost of keeping up with of state-of-the-art NMT technology standards is also high. For example, a current best-of-breed English to German NMT system requires tens of millions of training data segments, thousands of hours of GPU cycles, deep knowledge to tune and adjust model parameters and expertise to bring it all together. Just for one single system, estimated charges for training time on a public cloud infrastructure are around $9,000 and would take 40 days. These costs are likely to be higher if the developer does not have real expertise and is learning as they are working. These costs can be reduced substantially by moving to an on-premises training setup and by working with a foundation system that has been set up by experts.
4. NMT model development requires constant iteration and continuous experimentation with varying data sets and tuning strategies. Any model deployment carries a certain amount of uncertainty and outcomes cannot always be predicted upfront, so repeated and frequent updates should be expected. As a result, computing costs can rapidly escalate when using cloud infrastructure.

Given the buzz around NMT, many naïve practitioners jump into DIY (do-it-yourself) open-source options that are freely available, only to realize months or years later that they have nothing to show for their efforts.

The many challenges of working with open-source NMT are **covered here**. While it is possible to succeed with open-source NMT, a sustained research and production investment with very specialized human resources is required to have any meaningful chance of success.

As a technology provider focused on enterprise MT needs, SDL already provides existing adaptation capabilities, which range from:

- Customer-created dictionaries for instant self-service customization – suitable for specific terminology enforcement on top of a generic model.
- NMT model adaptation as a service, performed by the SDL MT R&D team.

This situation will now change and continue to evolve with an innovative new NMT adaptation solution **being introduced by SDL**, which is a hybrid of the MT vendor and DIY approach - a solution that provides the best of both worlds.



Adaptability - NMT Trainer

Provides the ability to train Neural engines on-premise without the need to share data with any third party.

The SDL NMT Trainer solution provides the following:

- Straightforward NMT model adaptation without requiring users to be data scientists or experts
- Foundational data provided in the Adaptable Language Pairs to expedite the development of robust and deployable systems quickly
- On-premise training that prevents any highly confidential training data that encapsulates customer interactions, information governance, product development and partner and employee communications ever leaving the enterprise
- Once created, the encrypted adapted models can be easily implemented by SDL in an on-premise or cloud deployment with no possibility of data leakage
- Multiple use cases and optimizations can be developed on a single language pair and customers can continuously re-train and adjust their models as data becomes available or new use cases identified
- A pricing model that encourages and supports continuous improvement and experimentation on existing models, allowing for many more use cases to be deployed on the same language combination

The initial release of the SDL On-premise Trainer is the foundation of an ever-adapting machine translation solution that will grow in capability and continue to evolve with additional new features.

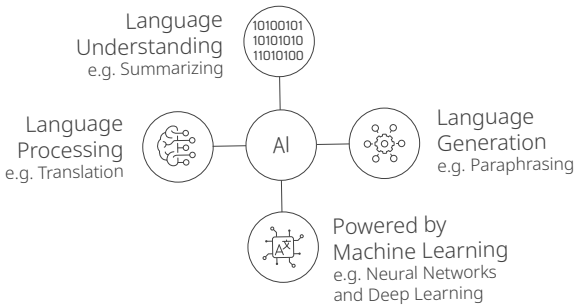
The promise of the market's best AI solutions is to continuously learn and improve with informed and structured human feedback and SDL technology is architected to evolve and improve with this human feedback. While generic MT suffices for many users who need a basic understanding of foreign language content, global enterprises need MT solutions that perform optimally on critical terminology and meet the enterprise's linguistic requirements. The preference is for a solution that allows a customer to produce high-quality adaptations with minimal effort in as short a time as possible, and thus make increasing volumes of critical digital experience content multilingual.

Looking Beyond MT

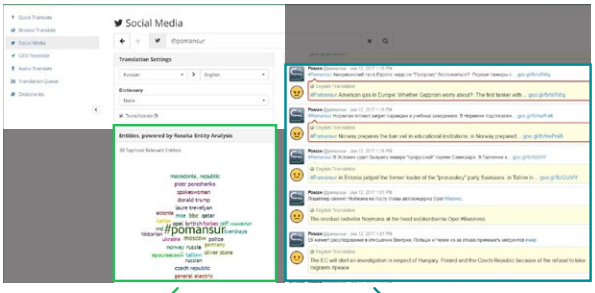
As we look from a larger perspective at how MT is being used by the modern enterprise, we see that it is increasingly necessary for data flowing through MT to be linked closely to other high-value NLP capabilities. These additional capabilities are needed to extract relevance and identify the most salient elements from the large volumes of data that tends to flow through MT.

We at SDL call these affiliated technologies Linguistic AI. Some key functionality is closely related to natural language understanding and natural language generation (NLG), where an NLG tool like SDL Content Assistant can analyze complex documents, provide a summary and generate "snackable" content to help marketers and content teams quickly and easily create content and understand a document's context.

What is Linguistic AI?



An example of how this may work is shown below where analyzing a foreign language Twitter stream can be usefully summarized while also identifying key topics and sentiment. As a result, thousands of customer social media comments can be both translated and analyzed to rapidly show internal marketing and product management teams the source and drivers of positive and negative sentiments.



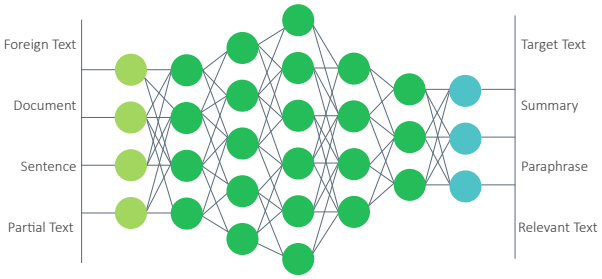
Entity word cloud extracted from Twitter feed (real time) in native language (Russian)

Social media feed translated in real time from Russian to English; Addition of sentiment analysis and sentiment intensity

These affiliated Linguistic AI tools may add further value to the relevance extraction process being applied to large volumes of content that is being translated.

Linguistic AI has many other relevant uses including:

- Summarization and classification
- Content paraphrasing
- Sentiment and emotion detection
- Content and topic extraction
- Content analysis
- Domain and language detection
- Content generation assistance

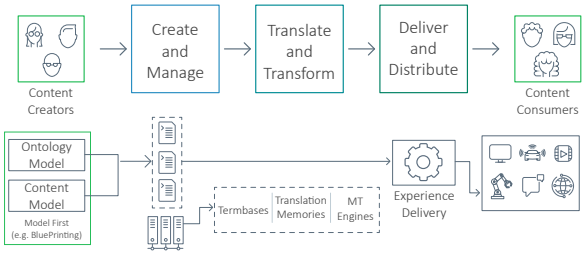


Linguistic AI Beyond Machine Translation

Linguistic AI helps the enterprise to understand and identify actionable information in a variety of market intelligence, customer monitoring and general surveillance settings.

Finally, it should be understood that MT is just one of many components in the overall content lifecycle and management process. Optimizing the content creation to content delivery process enhances an enterprise's ability to swiftly respond in this era of digital transformation.

SDL Innovation
Machine Learning through the ENTIRE Content Management Process



MT Technology Evaluation Checklist

The evaluation and selection of an MT system and preferred vendor is a much more nuanced and complex process than choosing an MT system based on BLEU/LEPOR scores on news domain data. There are many factors to consider and often the importance of these factors can vary based on the specific use case for the MT application.

An MT system used to manage critical daily communications between employees and global business partners and monitor urgent global customer support issues may have the following weighting of system requirements in order of importance:

- Data security and privacy
- Embedded into communication and collaboration IT infrastructure
- Adaptation to critical business and product terminology
- Basic linguistic quality of MT output
- Deployment flexibility

eDiscovery, Information Governance and Data Analytics

An MT system used for broad eDiscovery purposes demands a different set of requirements, such as:

- Data ingestion and acquisition capabilities
- Integration into analytical platforms for the core subject domain
- Ability to rapidly tune and modify the MT system to understand key terminology
- Ability to train MT systems with confidential data that cannot be sent to a vendor for processing
- Systems integration capabilities
- Specialized capabilities to handle Named Entities (NER)

Customer Service and Support

An MT system used to manage critical day-to-day communications between customers and provide relevant customer experience-related content may have the following weighting of system requirements in order of importance:

- Adaptation to critical business and product terminology
- Integration with support and content dissemination IT infrastructure
- Data security and privacy
- Embedded into communication and collaboration IT infrastructure
- Expert consulting services
- Consequently, we see that for several high-impact use cases the key requirement is not MT output linguistic quality, rather it is more likely to be one or a combination of the following factors:
- Data security and privacy
- Ability to incorporate and adapt to enterprise terminology
- Integrate directly into enterprise IT infrastructure to maximize efficiency and ease
- Solution deployment flexibility
- Availability of expert MT and linguistic resources

To summarize, the ability to control key elements of an MT solution and the evaluation process is more important than quality scores.

The factors that matter can vary greatly and buyers should look for MT vendors that can provide a broad range of capabilities and have a clear enterprise focus. The following checklist contains key factors that could and should be considered, scored and weighted.

Evaluation Factor	SDL	MT-A	MT-B
Language Coverage	•		
Data Security and Privacy Rating	•		
Data Security Certifications	•		
Secure Communication in Transit (TLS)	•		
Overall System Adaptation Options			
User-Managed Adaptation Options	•		
Secure Use of Private Data	•		
Secure On-premise Training with Confidential Data	•		
Ongoing Use and Maintenance of Custom MT Systems	•		
Expert MT Engineering Consulting Services	•		
Expert Linguistic Steering Consulting Services	•		
Deployment Flexibility			
Secure Cloud Options	•		
On-premise	•		
Edge / Small Footprint	•		
Combined (Edge – Cloud)	•		
Deployment Cost			
On-premise Options – Hardware Requirements	•		
Cloud Options	•		
Output Linguistic Quality			
3rd Party News Domain			
Enterprise Domain	•		
Use Case Specific	•		
MT System Manageability			
Ease of Deployment (No Professional Services Required) • Packaged installers • Installation utilities • No additional database installations required	•		
Options for Operating Systems	•		
System Administration Console Capabilities and Interface	•		
Ongoing System Management Requirements	•		

Integration Flexibility	SDL	MT-A	MT-B
API Features and Capabilities	•		
Integration into Communication and Office Productivity Apps	•		
Key Analytics Applications Integration and Plug-ins • TMS and TM • Social media • eCommerce • Data analytics • ASR and OCR	•		
Vendor Reputation and Profile			
Ability to Implement	•		
References	•		
Product Roadmap and Vision	•		
Use Case Suitability			
Experience and Customer Profiles	•		
Understanding of Key Use Case Requirements	•		
User Management and Controls	•		
Varied User Roles and Permissions	•		
User Authentication • None • App managed authentication • Enterprise active directory authentication (ADFS)	•		
MT System Performance			
Scalability and Robustness (Billions of words/hour)	•		
Production Throughput and Latency	•		
Controls to Manage Peak Loads	•		
Systems Integration Capabilities	•		

SDL, part of RWS Holdings plc, is the intelligent language and content company. Our purpose is to enable global understanding, allowing organizations to communicate with their audiences worldwide, whatever the language, channel or touchpoint. We work with over 4,500 enterprise customers including 90 of the world's top brands and the majority of the largest companies in our target sectors. We help our customers overcome their content challenges of volume, velocity, quality, fragmentation, compliance and understanding through our unique combination of language services, language technologies and content technologies.

Are you in the know? Find out why the top global companies work with and trust **sdl.com**.
Follow us on **Twitter**, **LinkedIn** and **Facebook**.

Copyright © 2021 SDL Limited. All Rights Reserved. The SDL name and logo, and SDL product and service names are trademarks of SDL Limited (for and on behalf of its subsidiaries and affiliates), some of which may be registered. Other company, product or service names are the property of their respective holders.