

## VENDOR NEEDS AND STRATEGIES

---

### Overcoming the MT Quality Impasse

---

Steve McClure

Mary Flanagan (contractor)

---

#### IDC OPINION

---

The difficulty of measuring the quality of automatic language translation systems (known as "machine translation" [MT]) has been an obstacle to widespread adoption. With systematic benchmark testing, categorization of errors, and effective dictionary customization, MT technology can yield significant cost and time savings, as well as improved consistency in translations. IDC makes the following observations about the MT market:

- ☒ MT quality must be assessed in the context of each user's application. For example, using MT for a chat or instant messaging application is completely different from using MT to translate manufacturing assembly instructions.
  - ☒ SYSTRAN has evolved a process for enhancing MT quality for an individual customer. This process has been validated with actual customers, such as Ford Motor.
  - ☒ SYSTRAN has also developed the SYSTRAN Review Manager (SRM), which helps the customer to manage the MT quality process by allowing them to change vocabulary and linguistic rules. This tool represents an important advance in MT, both technologically and philosophically. Users have never before had the power to modify linguistic rules through an intuitive, interactive process.
  - ☒ By opening up rule modification, SYSTRAN takes a risk, but one that will almost certainly pay off. Engaging users in the process of improving MT is the surest path to increased acceptance and understanding of the technology.
-



## IN THIS STUDY

In this IDC study, we discuss the efforts of SYSTRAN to address the issue of machine translation (MT) quality. MT technology faces several important obstacles to broader use in business applications. Most important among them is the issue of quality. Potential users perceive MT quality as uncertain and difficult to improve. All MT implementations require customization, and the scope of this task can be difficult to quantify. Such uncertainties can make ROI calculation difficult or impossible. But despite the perceived intangibility of MT, translation results can be measured and managed effectively. The key is targeted customization of the MT system, with ongoing benchmark testing to ensure that results continue to meet the customer's needs.

To address the quality impasse, SYSTRAN has developed the SYSTRAN Review Manager (SRM), a suite of quality management tools that address the measurement, testing, and customization tasks from a single user-friendly interface. Deployed successfully at Ford Motor Co., the SRM shows us the way to make MT quantifiable and viable for enterprise applications.

## SITUATION OVERVIEW

### INTRODUCTION

Despite the growing business imperative to produce and use multilingual business materials, automatic translation technology (known as MT) faces many obstacles to widespread adoption. Three of the most important are:

- ☒ **Determining return on investment (ROI).** There is currently no reliable method for calculating ROI. A useful formula for determining ROI would have to account for a broad range of factors, including language pairs, turnaround time, integration with existing applications, volume, customization, and support.
- ☒ **Evaluation challenges.** Translation technology is unfamiliar to most businesses. Comparing among competing products is difficult because it requires substantial knowledge of language technologies as well as a clear understanding of the specific translation requirements. MT products are completely nonstandardized, making feature comparisons difficult and further complicating the difficulty of evaluation.
- ☒ **Uncertainty about quality.** The difficulty of assessing and improving translation quality is MT's most intractable problem. Translation quality is inherently subjective and therefore difficult to measure. This is also true of human translation. A text given to three different professional translators will yield three different results. Even if all three are of high quality and accuracy, the subjective nature of human language almost guarantees that there will be differences in interpretation, word choice, and style. It can also be difficult to quantify the effort required to improve translation quality. Tracking ongoing quality levels is also a challenge.

### THE MT QUALITY IMPASSE

There is an enduring perception that MT is not yet "good enough" for commercial use. The irony is that MT is better today than it ever has been, and it is in broader use. For many high-volume translation applications, the quality of translation is sufficient to allow understanding of the text. In addition, some MT systems, notably SYSTRAN,

have introduced powerful linguistic and integration tools that have increased the user's ability to customize the MT system.

Still, the problem of assuring and measuring MT quality remains a serious concern among potential users. Many of these concerns are well founded, because it is difficult to track and quantify MT quality. There are a variety of factors at the root of the difficulty, which are discussed in the following sections.

#### MACHINE TRANSLATION OUTPUT IS NOT EASILY PREDICTABLE

MT systems work with natural language — a data set that is infinitely varying, ambiguous, and structurally complex. To translate adequately, an MT system must encode knowledge of hundreds of syntactic patterns, variations, and exceptions, as well as relationships among these patterns. It must include ever-changing vocabulary and specific semantic knowledge about the usage patterns of tens of thousands of words. It must accurately identify the parts of speech and grammatical characteristics of words which may, in different contexts, be nouns, verbs, or adjectives, each having many possible translations. Translation also requires a vast store of knowledge about the world, the intent of the communication, and the subject matter.

A human translator prioritizes and selectively applies linguistic rules based on this knowledge. MT software, unless explicitly coded for each possibility, cannot. Thus, MT will never attain the overall quality of human translation. The primary advantages of MT over human translation are speed, cost, and consistency. An MT system gets a great deal more translation done than is possible manually, and MT can deliver translations instantly for time-sensitive content. When a term is entered in an MT dictionary, it will translate it the same way every time, unlike human translators who may choose different translations at different times.

#### QUALITY METRICS DEPEND ON THE INPUT TEXT AND THE LEVEL OF CUSTOMIZATION

Potential users want quality metrics that are objective, absolute, and easily compared among competing products. However, translation quality, whether human or software generated, is difficult to quantify. Counting the number of errors in a translated sentence is not revealing because languages do not correspond on a word-for-word basis. An incorrect analysis of one word in the source language, for example, could lead to incorrect translation of several words in the target language. In addition, many errors made by MT systems cause subsequent errors within the sentence. Different systems, and for that matter, different human translators, can produce intelligible, accurate, but different translations of the same sentence. Therefore, for any input sentence, there is no single, ideal output sentence. Finally, some errors are more serious than others, so all errors should not be assigned the same importance.

#### NO STANDARDS GOVERN MT SYSTEMS

Despite the decades of research and development that went into today's MT systems, the industry is still immature. MT systems grew up in very different ways, with many originating with academic research projects or government-funded initiatives. As a result, there are no accepted standards for how MT systems store or process data or what results they produce. Without a standard to measure against, each system vendor is left to make their own claims, which are not directly comparable with the claims of competitors.

#### EVALUATION IS NOT OBJECTIVE

One evaluator might rate a translation as intelligible, while another may not. The judgment of translation understandability is an inherently subjective task that can be

affected by factors, including the evaluator's subject knowledge, language facility, reading comprehension, translation experience, and attentiveness.

#### A SUCCESSFUL STRATEGY FOR THE MT QUALITY IMPASSE

Many potential users give up when faced with the challenges of evaluating, enhancing, and implementing MT. MT vendors recognize the risks, and most have responded by working to improve basic translation quality to increase acceptance. But for most applications, improved translations are not enough. Adopters of MT need comprehensive, easy-to-use tools for measuring the quality of their translations, enhancing dictionaries, and verifying the results. The tools must be accessible to nondevelopers who know the languages and the business terminology for their company. Among the handful of commercial MT systems available today, only SYSTRAN has tackled the quality issue effectively.

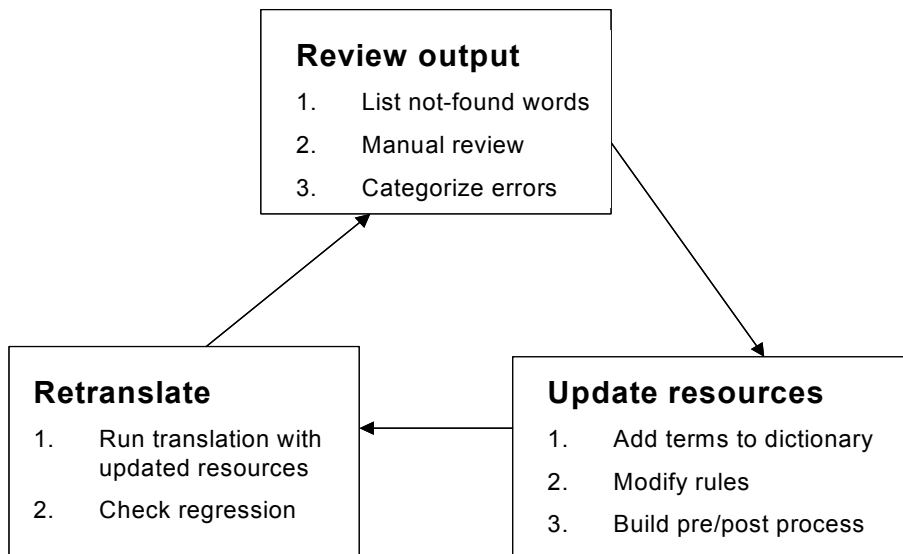
SYSTRAN is unarguably the best-known and most comprehensive MT system in the world, having been in continuous development for more than 35 years. SYSTRAN offers 36 language pairs and has the largest dictionaries of any MT system. The company has taken a pragmatic approach, developing a suite of quality measurement and enhancement tools that offer a far more concrete solution to the quality question than esoteric measures of improvements in basic translation quality.

#### THE MT QUALITY ENHANCEMENT PROCESS

Making an MT system work for a particular application is a *process*, not a quick fix. Improving MT is a cyclic process beginning with review of a translation, update of dictionaries and other linguistic resources, and retranslation to validate the effects. In the SYSTRAN system, the SRM acts as a coordinator, managing access to different customization resources and tracking quality.

**FIGURE 1**

#### MT QUALITY ENHANCEMENT PROCESS



Source: IDC, 2003

With potentially thousands of dictionary changes, numerous rule modifications, and changing text, it is a challenge to track customization activities and measure results.

The SRM integrates the three steps into a single-process management program with links to the user dictionary, the source and target texts, benchmark files, and interactive translation testing. In addition, the SRM categorizes errors, assigns levels of severity, and keeps track of statistics on the rates of various error types. It can be configured as a Web-based application for single or multiple users. In the latter case, reviewers in different locations can access translations, provide feedback, update dictionaries, and even store their own variant translations for a particular word or phrase. For multinational companies, the SRM allows easy cooperation between sites where different language abilities reside. Some additional benefits of the SRM are:

- Demonstrable method of quantifying MT results
- Increased user autonomy in the enhancement process
- Reduction in the need for continuing customization services from a specialized provider
- Leveraging of the company's own multilingual resources regardless of location
- Increased QA productivity and deeper user engagement in the quality review process
- Improved efficiency in managing translation projects

#### STEP 1: REVIEW OUTPUT USING THE SRM

During this phase, the SRM functions as an interactive editor, presenting the reviewer with each translation unit in the translated text. The user can modify the translation if it is not acceptable. These modifications are recorded as new entries in the User Dictionary. In the soon-to-be-released version 5.0, the SRM can automatically determine grammatical information, such as part-of-speech and inflection patterns, and enter that information into the new dictionary record for the term or phrase. This function is known as "Intuitive Coding." With Intuitive Coding, people with language and subject knowledge can encode the dictionary without any special expertise in linguistics or programming. The reviewer can also view listings of words that were found in the text, but have not yet been entered in the dictionary. These listings can be entered directly into the User Dictionary from the SRM. The reviewer supplies the translation, and the Intuitive Coding functionality supplies grammatical information.

#### STEP 2: UPDATE RESOURCES AND ENHANCE SOURCE TEXT

After the review process is complete, the dictionaries are saved, and the document can be retranslated. Reviewers can also open the dictionary records directly and modify or refine the translations or grammatical tags for an entry.

Enhancing the source text is equally important to dictionary building for quality assurance. Translation results tend to be better when the source text is modified to simplify word order and shorten lengthy sentences. SYSTRAN is developing an interactive linguistic tool that allows reviewers to modify the actual translation rules used by the translation engine. Combined with the SRM, the SYSTRAN Translation Workbench is an interactive XML-based editing tool that incorporates the reviewer's changes as rule modifications.

Once it is released, this tool will represent an important advance in MT, both technologically and philosophically. Users have never before had the power to modify

linguistic rules through an intuitive, interactive process. Rule access was provided once before in the translation engine developed in the 1990s by Globalink. Code-named "Barcelona," that system was subsequently sold to Lernout & Hauspie and Bowne Global Solutions. The rule language of Barcelona, though powerful, was extremely complex, requiring a great deal of skill in linguistic notation, programming, and languages to use it effectively. In most MT systems, linguistic rules are not even accessible to the user because they are part of the source code.

Perhaps most importantly, the coming release of the SYSTRAN Translation Workbench represents a shift in the attitude of MT developers toward users. MT systems are extremely complex, and developers have always taken pains to protect the user from making naïve changes to the system that could have serious consequences for other contexts. This attitude has been a source of frustration to more sophisticated MT users, who eventually reach a wall on quality improvements after building their dictionaries. By opening up rule modification, SYSTRAN takes a risk, but one that will almost certainly pay off. Engaging users in the process of improving MT is the surest path to increased acceptance and understanding of the technology.

#### STEP 3: RETRANSLATE AND VALIDATE

Once the changes to the system are saved, the reviewer can retranslate the text to verify that the new entries are in effect. It is important at this stage to check for regressions. Regressions occur commonly in MT output. They can sometimes originate with an incorrectly coded dictionary entry. For example, a user might supply a translation that is correct in the context of one sentence, but incorrect in another context.

The SRM manages regressions with a color-coding system that shows what portions of the text have changed since the last time it was translated. This feature reduces the amount of time spent on reading and comparing the previous translation with the new version by highlighting the areas for focus.

#### **SIGNIFICANCE OF THE SRM**

The SRM will benefit SYSTRAN's customers by improving their understanding of translation quality and the process for improving it. The SRM also has broader importance, in that it places far more control over the translation process in the hands of the user than ever before. This may spur changes to the way the MT industry and its customers view each other and lead to more successful implementations of MT.

---

#### CASE STUDY: SYSTRAN AND FORD MOTOR

Ford Motor Co. is an example of the MT imperative on a grand scale. Ford has manufacturing facilities in Germany, Spain, Belgium, Mexico, and Brazil, where workers assemble vehicles using instructions in their local language. However, all of the instructions are originally created in the United States in English. A single car line can have assembly instructions with as many as 300,000 sentences. Moreover, the instructions undergo frequent changes during the production cycle, requiring quick retranslation and distribution. For such a massive translation problem, MT is the only viable solution.

Ford engineers prepare the assembly instructions using a standardized language. This language has a limited range of syntactic patterns and vocabulary to reduce the possibility of ambiguity. When assembly instructions are prepared, each standard language sentence is stored as a record in an Oracle database. Ford developed its own artificial intelligence system to check the sentences for conformity to the Standard Language rules. When Ford needs to translate the instructions for a

particular vehicle and language, the appropriate records are sent to the SYSTRAN MT system, where they are translated using Ford's customized dictionaries and rewritten to the database. The translated instructions can then be sent directly to the PCs of Ford workers at manufacturing sites worldwide. Currently, Ford is producing MTs in four languages: German, Dutch, Spanish, and Brazilian Portuguese. The database, SYSTRAN system, and customized dictionaries are integrated into Ford's Global Study Process Allocation Process (GSPAS), a system for managing labor and manufacturing costs for Ford plants worldwide.

### **UNIQUE CHALLENGES**

Every MT implementation involves a unique set of customization challenges related to the nature of the text and the intended audience. At Ford, some of these challenges were:

- ☒ The texts contain numerous long noun phrases (e.g., insulation assembly body pillar), which must be recorded in the Ford user dictionary to ensure an accurate translation.
- ☒ All Standard Language sentences are written in Imperative form. Declarative sentences are the most prevalent type in most English texts, so grammatical coverage of imperatives tends to be less robust.
- ☒ Standard Language uses modification rules that are different from the rules for English. Modifying words can be placed after a noun, instead of before it. For example, the phrase "body panel large" is allowable in Standard Language, even though it is grammatically incorrect in English.
- ☒ Ford uses a specialized vocabulary. Some of the vocabulary is common to automotive manufacturing in general, but some terms can be specific to the specific plant or manufacturing team. Standard language contains 2,500 Ford-specific terms, 13,000 noun phrases, and over 1,000 abbreviations and acronyms. Ford uses an artificial intelligence system to review its assembly instructions and ensure they conform to the Standard Language rules.
- ☒ Spelling variants are common. The acronym for "antilock brakes" may be written as either ABS or A.B.S.
- ☒ Writers can insert free-form comments that do not conform to the Standard Language rules.
- ☒ Ford's bilingual engineers do not have the time to review translation results.
- ☒ Standard Language is usually written with no punctuation. MT systems are sentence based, and they rely on proper punctuation to help segment sentences, clauses, and lists.
- ☒ Standard Language is always evolving. The MT system and its dictionaries need updating to account for the changes in Standard Language.

### **MT INTEGRATION AT FORD**

Ford and SYSTRAN collaborated successfully to address these challenges, integrating SYSTRAN into the GSPAS system in 1998. Today, the system is in use at Ford's worldwide manufacturing plants.

SYSTRAN analyzed Ford's texts to identify frequently occurring technical terminology and built a custom dictionary for the application. It was also necessary to map

abbreviations to full words (e.g., [ASSY — ASSEMBLY]). SYSTRAN modified its translation system to account for modifiers that occur after the noun. Dictionary development is only one part of linguistic customization. To customize the rules of the translation system, SYSTRAN uses an XML-based "style sheet" that allows users to select from configurable rule categories. The categories of errors can be tabulated after the review is complete, offering insight into the nature and frequency of problems in the translation.

After initial tests, it was clear that some preprocessing of the assembly instructions would help translation quality, especially with embedded free-form comments and titles, neither of which conform to the Standard Language syntax. In addition, inserting articles (e.g., the) before nouns would help the MT system to identify the correct part of speech. For some languages, these problems are being addressed by automatically preprocessing the text prior to translation.

Ford also identified dialect and text size differences as important areas for quality enhancement. Many languages have variant dialects, though the differences in speech are usually far more extensive than in written English. For example, a coastal Maine resident and someone from the deep South might have difficulty understanding each other's speech. But in written form, their language is very much the same. The same principle applies with translation. In Spanish especially, there are numerous dialects. Although the differences are more prominent in speech than in writing, there are nonetheless some terminology issues among Spanish dialects. The quantity of text for any given message varies depending on the languages involved. For English to Spanish translation, for example, translations are generally 15–20% longer in Spanish than in English. This has implications for how the text is displayed and the size of the text window in the user's application.

## **FUTURE OUTLOOK**

The use of SYSTRAN has helped Ford to translate its large volumes of assembly instructions into four languages. More than 1 million records have been translated. Ford has been able to deliver an accuracy rate of 90% for English/German translations. Ford deployed a Web-based customer dictionary tool in 2002 that allows engineers to introduce new dictionary entries and corrections to translation errors. Modifications to the Standard Language have been introduced as a result of translation feedback.

## **ESSENTIAL GUIDANCE**

The MT quality impasse can be overcome with customization, ongoing error tracking, and testing against benchmark files. This process can be intimidating to new users who are unfamiliar with MT technology. To deploy MT successfully, the vendors must provide guidance and support to the user until sufficient knowledge is built up within the organization to manage translation quality independently.

When it is effectively customized and tested, MT does produce cost and time savings. Ridding potential users of the notion that MT is a "plug and play" solution is perhaps the MT industry's most important educational objective. The SYSTRAN implementation at Ford Motor provides an excellent case study of how MT, when properly customized, can solve a critical, large-scale translation problem. Other MT users and vendors would do well to follow Ford's example.

## LEARN MORE

---

### RELATED RESEARCH

- ☒ *Machine Translation Engines: An Evaluation of Output Quality* (IDC #22722, June 2000)
- ☒ *MT and TM: ESteam's Winning Solution* (IDC #28367, November 2002)
- ☒ *Worldwide Globalization/Translation Software Market, 2002* (IDC #28166, November 2002)
- ☒ *SYSTRAN and the Reinvention of MT* (IDC #26459, January 2002)

---

### COPYRIGHT NOTICE

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit [www.idc.com](http://www.idc.com) to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit [www.idc.com/offices](http://www.idc.com/offices). Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or [sales@idc.com](mailto:sales@idc.com) for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2003 IDC. Reproduction is forbidden unless authorized. All rights reserved.

---

**Published Under Services:** Globalization and Translation Software; Globalization Solutions