

Applying Automated Language Translation at a Global Enterprise Level

Nestor Rychtyckyj, Craig Plesco

■ *We have been utilizing natural language processing (NLP) and machine translation (MT) at Ford for more than 10 years for the automated translation of process build instructions from English to other languages in support of Ford's assembly plants in non-English speaking countries. This project has continued to evolve with the addition of new languages and improvements to the translation process. However, we discovered that there was a large demand for automated language translation across all of Ford Motor Company and we decided to expand the scope of our project to address these requirements. This article will describe our efforts to meet all of Ford's internal translation requirements with AI and MT technology and focus on the challenges and lessons that we learned from applying advanced technology across an entire corporation.*

We have been applying AI and machine-translation (MT) technology at Ford Motor Company since the late 1990s. Our initial goal was to utilize MT to translate vehicle build instructions from English to the native languages in the countries and regions where our assembly plants are located. The source text utilized a controlled language that we developed, called Standard Language, and we initially thought that applying MT technology would be a straightforward process. Controlled languages, such as Standard language, restrict the complexity and ambiguity of human languages by restricting syntax and terminology (Huijsen 1998). As such, they have been utilized in a number of different industrial applications (Godden 2000). However, there were many issues dealing with technical terminology, ungrammatical aspects of Standard Language, Ford-specific terminology, and the need to process uncontrolled text that needed to be addressed. We partnered with Systran Software Incorporated and with AppTek (now SAIC) to use their machine-translation technology and also incorporated natural language processing (NLP) algorithms within our artificial intelligence environment to analyze terminology and modify the source text to improve translation accuracy (Rychtyckyj 2007). The need to support manufacturing expansion in non-English speaking countries in Eastern Europe and Asia (such as in Russian and Chinese) led us to add additional language capability and to develop translation glossaries for all of the supported languages. The automated language translation for manufacturing work continues and will expand as Ford's global manufacturing footprint increases. However, the international growth within the company was not limited to manufacturing only and we found that there are many different groups within the company that need some type of machine-

translation solution. Therefore, in 2010 we deployed an internal web-based machine-translation solution that sought to leverage our work in manufacturing and make automated translation a reality for the entire company. In the following sections we will describe the process of delivering a new technology across an entire company and lessons that we learned.

Machine translation has become ubiquitous in the last few years. Since the advent of the first MT systems in the 1960s the technology has been supported by a few specialized vendors and the cost to develop machine-translation systems was significant. This situation has changed dramatically in the last few years as the introduction of statistical machine translation has decreased the development time and subsequently large companies such as Google and Microsoft have become heavily involved in machine translation. The main result of this is that most users have had some experience with the technology and will likely have some type of preconceived bias (either positive or negative) when they are introduced to it as part of their daily work.

Unfortunately, many users still treat machine translation as a “black box” technology and expect to receive high-quality translations suitable for their specific purposes (conversational, business unit jargon, and so on) given any sort of input without having to do any other work. Other users have had bad experiences and do not believe that machine translation can work well in any instance. A large part of our work is to educate and manage these user expectations so that they can use the technology effectively. For example, a very common request that we have is to translate screen headings and labels into another language as part of a conversion process. These headings are usually one or two word phrases that often contain abbreviations and acronyms. This type of translation is difficult for machine (and human) translation because there is very little context available and these phrases may be ambiguous in many cases without a detailed knowledge of the application. In these cases, it is critical that the users be aware of these limitations and have human posteditors available. It is always important to ensure that these translations and edits can be reused for other applications, as many business-specific phrases can be shared.

It is also difficult to determine how well the technology will work for a given application. In general, MT works better when the source text contains complete grammatical sentences that use correct punctuation and articles in front of nouns, but the performance may vary greatly. One common request for translation usage is with presentation material. In this case, the presentation writers are taught to remove extraneous text so the slide does

not look too crowded. However, this extraneous text is often needed to help the translation system develop more accurate translations. Therefore, it is always better to have the authors take these translation rules into account when they are creating the source materials.

There are many issues with applying new technologies at a global and/or enterprise level. These include things such as cost and maintainability, acceptance by the user community, the management of user expectations, corporate standards, and support. A perspective on applying soft computing technologies at Dow Chemical Company is presented by Kardon (2005). He points out that much of the skepticism and resistance to new technology is based on organizational and political issues where appropriate management support, lack of acceptance of the technology by the business user community, and low credibility are all barriers to acceptance of new technologies. Closer to the machine-translation realm there have been a number of applications of MT at an enterprise level at companies such as Paypal and Microsoft (Beregovaya and Yanishevsky 2010; Dixon 2010; Depraetere and Vazquez 2010). Each application is different in the manner that machine translation is applied. In some cases it may be applied to an existing human translation process and in other cases it may be applied to only one part of the company. Our work had a somewhat broader objective — we wanted to make translation available across the enterprise and then apply customization at a level that a particular user community required.

Our article is organized as follows: first we introduce and discuss the translation system that was developed initially to support manufacturing. We then discuss the process by which we extended our work to include translation for other users within the company. This discussion will cover both the technical and organizational challenges that need to be addressed when introducing a new technology. We focus on the application of linguistic and natural language processing that is used to enhance the quality of the text prior to translation and the processes that we developed to integrate new applications into the translation process. Our article concludes with a discussion of the current usage of our system, future plans, and the lessons that we learned along the way.

Application Description

As mentioned previously, we had already deployed an integrated MT system into Ford’s global process planning for vehicle assembly system, known as the Global Study Process Allocations System (GSPAS). This application utilizes NLP to “clean up” the source text, takes advantage of Ford-spe-

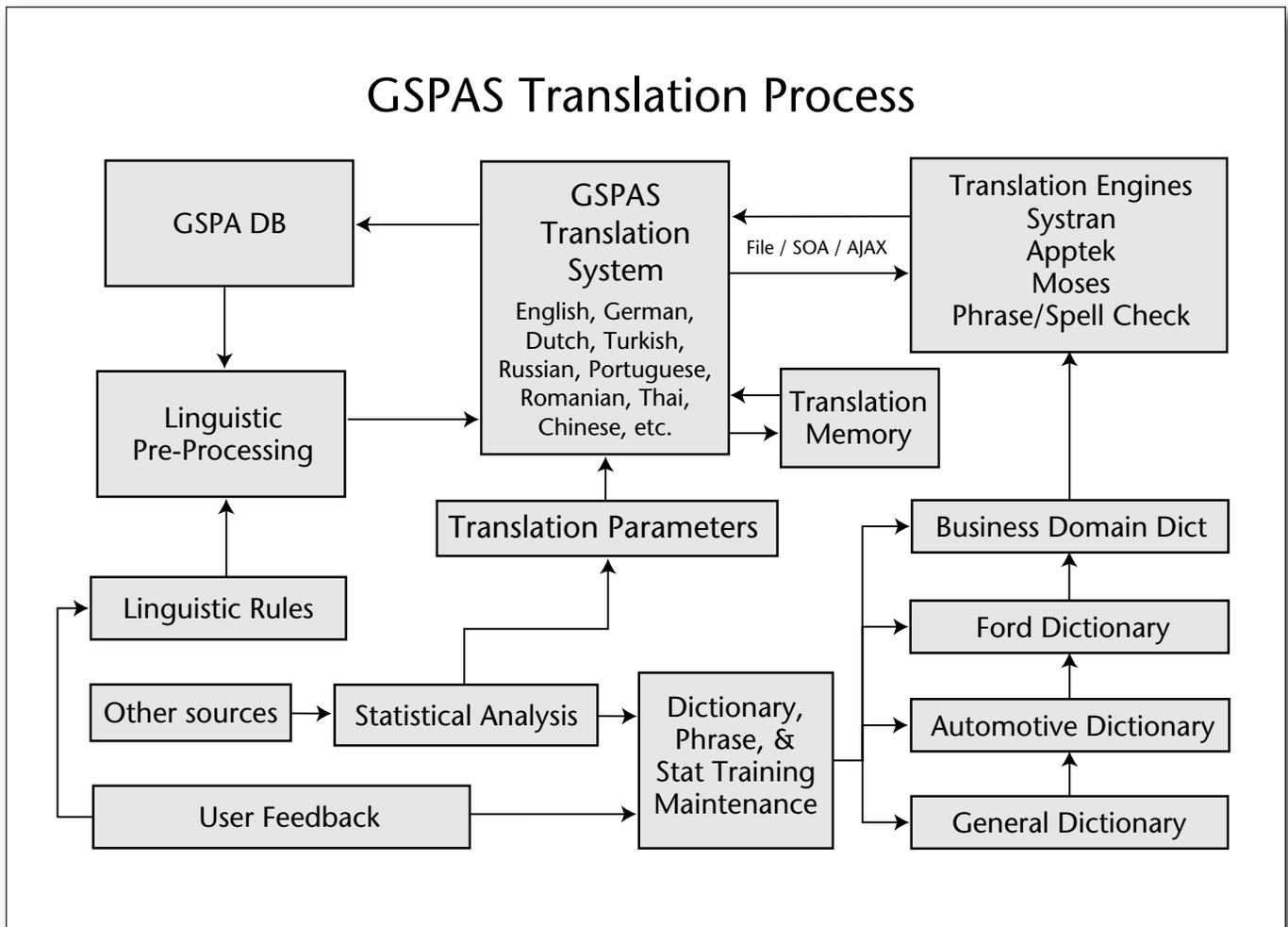


Figure 1. Automated Translation for Process Planning.

cific terminology translation glossaries, and works with customized Systran and SAIC (for Turkish) translation engines. It supports translation from English to seven other languages and has translated more than 23 million records. Translation accuracy can be as high as 94 percent as measured using the Society of Automotive Engineers (SAE) J2450 translation metric (Society of Automitice Engineers 2002). Figure 1 displays the architecture of the system.

The GSPAS translation process is completely automated; the input textual data is read in from an Oracle database and NLP linguistic algorithms are used to “clean up” the text prior to translation. The completed translations are then written back into the Oracle database. However, the process of validating and improving the translation quality requires manual input and user feedback. If a translation glossary is significantly changed, we will retranslate the existing translations in order to improve the quality. The system uses different translation dictionaries that contain Ford-specific,

automotive, general, and business domain terminology. The translation system first checks the Ford-specific dictionary, and if a term or phrase is not found it will continue to check the other dictionaries until a translation is found. Other translation resources, such as translation memories (text that has been previously translated) and internal dictionaries have been integrated into the system. These Ford-specific glossaries contain both single-word terms and entire phrases. Technical terminology that contains multiple words must be translated as a single entity. A translation glossary for a single language pair will usually contain at least 6000 to 7000 entries. These glossaries need to be created manually by somebody who is well versed in the language and in the domain.

The GSPAS system adequately supported our translation requirements for vehicle manufacturing, but it was soon readily apparent that there was a large need for this type of translation for users and applications outside of manufacturing. There were two separate types of requests for translation.

First, to provide some type of translation tool that could be used in an ad hoc manner to translate emails, documents, presentations, and similar types of internal documentation. Second, to provide a programmatic solution that could be used by other applications that required translation of textual data to integrate with our translation system.

We recognized that both approaches would need to be dealt with separately and our first priority was to set up an internal language translation website. This was done with the assistance of Systran by deploying a web-based Systran Enterprise Server internally and then adding the Ford-specific translation glossaries into the system to improve the translation accuracy. We then created an internal URL¹ and began the process of letting people know about our service. Since Ford is a global company with 166,000 employees located in more than 130 countries around the world, this was not a simple process. We started by contacting many of the local users that had expressed an interest in using translation technology, but also took advantage of the many communication facilities that are currently available. These included presentations and demonstrations at technical fairs and internal conferences, articles that appeared on internal websites, social networking sites (such as Yammer), and placement on some of the corporate internal websites. We also purposely disabled some of the security features in the software to let people use the tool without having to be registered, but also allowed people who did register to have access to some of the more advanced features. Certain groups wholeheartedly endorsed the tool and began collaborating with us to develop subject-area glossaries for their projects and made the tool part of their business process. Other users just took the opportunity to try the software and compare it to other translation tools that were available. We also found out that old-fashioned “word of mouth” advertising is still very effective in our digital social networking world.

From the beginning of this project we needed to demonstrate the need for an internal translation tool when there were free translation tools available externally. There were several important business requirements that could only be addressed by an internal solution to the translation problem. This includes security of the text and its translation, the capability to customize the translation for our needs, and the need to leverage the many translation resources that have been developed independently throughout the company.

We also realized that there were different levels of support that were needed by different users. They were grouped into the following three categories.

The first category included users who intermit-

tently needed to access the translation system for everyday tasks and did not use specialized terminology or require a high level of accuracy. These people were allowed to use the system with little or no assistance.

In the second category, other users needed to use the system as part of their everyday job; they needed specialized translation glossaries, translation memories, and other assistance to use the translation tool effectively. We worked with them to create and integrate these resources into the translation system.

In the third category, we also had a group of users that needed to integrate translation into their business processes and applications. This often required use of AI/NLP tools for analysis of the terminology and source text, preprocessing and “clean up” of the source text prior to translations, conversion of the software and database to support different character sets, and a programmatic interface into the translation engine. This level of support was needed for applications that needed to be converted for use in countries where English was not the native language and required substantial work. We created a Service-Oriented Architecture (SOA) that could be utilized by any application that required translation capability. The main difference between the two approaches lies in the development of a general web service call protocol that can be utilized by any application within Ford. This is represented in the top-left corner of the diagram where a request from an application is received and a completed translation is sent back. Figure 2 shows this in more detail.

The diagram in figure 2 demonstrates how we have built a general approach to integrating translation into different applications by providing translation as a service. The various translation glossaries can be controlled through the use of the translation profiles to ensure that each user utilizes the relevant glossary for his or her application. Information such as the languages to be used as well as specific linguistic information, such as “part of speech” tagging, is included in the call. The main advantage of this approach is that each application can utilize the translation system just by utilizing the generic interface that does the actual communication with the translation system.

The service is not published as a corporate openly published SOA module, but rather it is provided after prework is completed with the application and MT/AI specialists. If we agree that machine translation is appropriate, then our next step would be working with the application group to generate line-of-business-specific terminology glossaries, linguistic rules, and other configuration. The service is defined as XML over HTTP with input stream defining the following attributes:

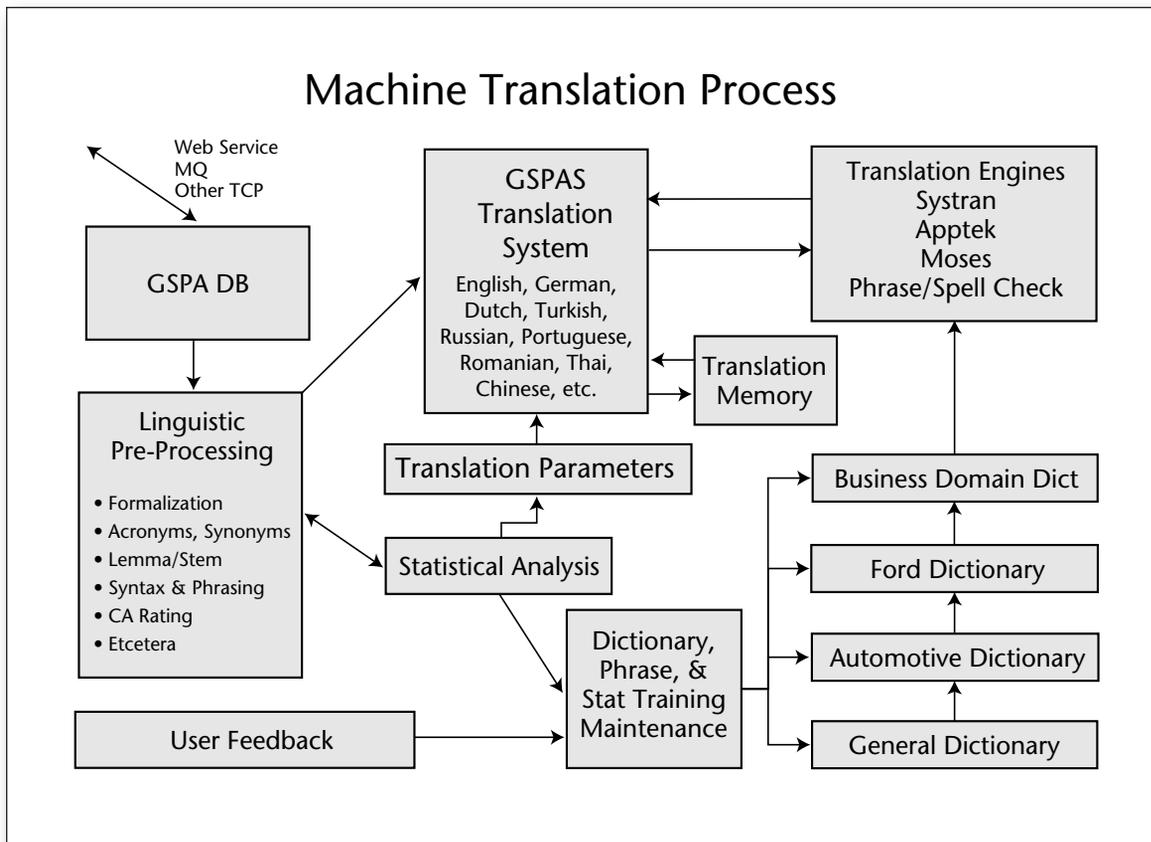


Figure 2. Machine Translation as a Service Diagram.

Application ID (predetermined for authorization)

Action (translate for now, but planned to support spelling checking and other functions)

Profile override (using the translation glossaries/rules other than default for application and language pairs)

Part of speech (if other than a complete sentence, we attempt to set parts of speech segments into prototypical pretranslated sentences for translation and extraction of the then translated subphrase)

Source language and country (to support language derivatives in different regions)

Target language and country (to support language derivatives in different regions)

Source text (the object that needs translation)

Special predetermined linguistic processing requests (such as pretranslation formalization and shortening length of returned translations)

Tracking key (identifies application specific use for phrase)

Based on previously agreed rules specific to an application and its supported language pairs, and included directives in the input XML, we preprocess and then translate the source text. Preprocessing can include formalization of the text, placement of sentence fragments into prototypical

sentences to enhance translation context, expanding synonyms and acronyms, stemming nouns, and other requirements of the application and the translation service. The text can be formalized by applying specific linguistic preprocessing rules (as described in the next section) and by providing additional information, such as “part of speech” and parsing information that can increase the accuracy of the translation. The output XML provides the following 7 attributes:

Key (other internal index to the request for metrics and error tracking)

Status (success or error indication)

Message (further details on status)

Confidence rating (our percent confidence of translation based on metrics including previous human translation — a high percentage, and phrase fragmentation — a low percentage)

Source text (source untranslated text for reference)

Translated text (the object that is translated)

Tracking key (returned verbatim from request so that client application can apply translation to reference record)

Any linguistic preprocessing that needs to occur is also included as part of the service and can be processed before the call is made to the translation

system. Multiple translations can occur in one XML request or reply stream for performance. On our side this is handled as serialized translation. The following section will describe the linguistic preprocessing in more detail.

Uses of AI Technology

We have partnered with Systran Software Inc. and SAIC to help with the translation capability. Machine translation has evolved significantly in the last few years and both systems utilize a hybrid approach that incorporates both rule-based and statistical approaches. The internal workings of the translation engines are mostly a “black box” to us, but we have developed and deployed AI-based approaches to help us analyze and understand the terminology and source text. The details of the internal machine-translation processing can be found for Systran (Surcin, Lange, and Senellart 2007) and Apptek (now SAIC) (Matusov and Kopru 2010). The results of this analysis are used to develop translation glossaries for each language that needs to be translated and helps us to manage and make the most of human translator contributions. This is done by using both home-grown NLP algorithms as well as available NLP tools to identify terminology within the source text that needs to be added into the translation glossary and to predict the accuracy of the translation. It also helps us identify terms (acronyms, abbreviations, slang, misspellings) that will need to be addressed prior to being sent to the translator. We also utilize a parser to identify parts of speech in the source text and this information is utilized to improve the translation.

Another task for AI was the development of a linguistic preprocessing algorithm. This was motivated by the realization that translation accuracy can be greatly improved by writing source text that conforms to good grammar rules. These rules can be summarized as follows: (1) Sentences should be written in a simple grammatical form. (2) Proper punctuation and capitalization need to be used. (3) Misspellings need to be corrected. (4) Acronyms and abbreviations need to be checked and replaced if necessary. (5) Nouns and compound nouns should be preceded by an article. (6) Shorter sentences are preferable. (7) Active voice should be used when possible. (8) Comments (separated by parentheses) in a sentence need to be translated separately.

Some of these rules (7 and 8) are enforced by the AI system when using Standard Language, but many translations do not use a controlled language, so our AI linguistic preprocessor tries to clean up the text as much as possible by applying these rules. The algorithm for the preprocessor is shown below. Since human language can often be

ambiguous we also try to account for those instances where the “clean up” leads to a sentence that is less understandable than the original. In these cases, the algorithm will revert back to the original sentence.

Start:

Read in input text.

Parse input text and create parse tree.

Check terms against automotive knowledge-base and replace acronyms and abbreviations with correct terms.

Identify noun phrases and place appropriate article in front of valid noun phrases.

Identify comments and place appropriate tags around the comments.

Check modified sentence against linguistic rules and revert back to original if violation is found.

Encode text to be translated and send to translation system

Our technical terminology raised the challenge of defining translation equivalents for the well-defined terms of Standard Language. There are many terms that describe automotive processes and product parts that are only utilized within our company. These terms include acronyms, abbreviations, company locations, and other terms that cannot be translated by someone who is not internally familiar with our company. It was found that many of the terms were not understood by all of our people, as they may only be used within one department in a plant. These terms all have to be identified and translated manually so they can be added into the Systran dictionaries correctly. Problems were also caused by entries (for example, “shotgun”) that are used informally to describe tools or equipment at the plant. Many other people may be unaware of what such a term represents, and a literal translation of “shotgun” would make no semantic sense in German or Spanish. Technical glossaries, such as those published by the Society of Automotive Engineers, are very useful in some cases, but they do not always contain a complete list of terms and can become dated and obsolete due to the rapid pace of technological progress.

Another issue with Standard Language arises with multiple spellings and misspellings of various terms. For example, some process writers would use acronyms without periods (ABS) and others would add periods to the term (A.B.S.).

The verbs in Standard Language are defined very concisely and unambiguously to represent a single particular action. It was not always possible to translate these verbs into other languages on a one-to-one basis to preserve their consistent meanings. The translation was accomplished only after spending considerable time on redefining their meanings in English and then translating the verb based on the most common usage in the target language. In some cases, one single English verb

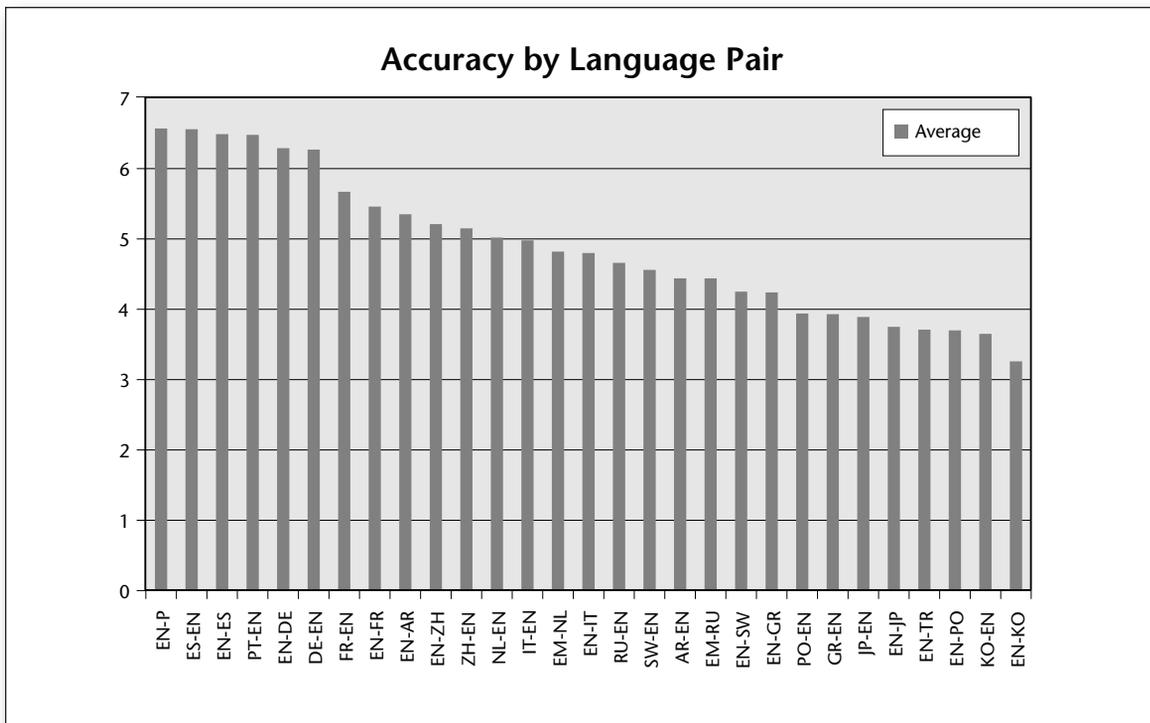


Figure 3. Survey Responses for Translation Accuracy.

would have multiple translations based on its context or the object that it was acting upon. Another problem arose with the use of compound verbs, which are a creation of Standard Language. Compound verbs (such as “press-and-hold”) are created to describe two actions that often occur together. Their usage makes it simpler for the process writers but causes complications in translations, as we are creating a new word in another language. Other languages cannot always represent the meaning of combined actions in the same way that they are described in English. The same issue occurs with certain acronyms and abbreviations, which may need to be handled differently in other languages. The entire issue of defining an equivalent Standard Language lexicon for each of the target languages required considerable effort and is not entirely completed.

We are following a similar approach for terminology that is not part of a controlled language. The obvious issue here is that the terminology and syntax are not controlled and the linguistic processing needs to be much more robust than with a controlled language. There are also many differences in text quality between various applications. In some cases, we need to deal with very cryptic unstructured text with many abbreviations and acronyms (an example would be “CHK. FRT WHL FOR EXS VIB”) while other applications have text that is written in fairly well-structured English. All of this source text data needs to be analyzed and the terminology that cannot be translated using

the generic system needs to be identified and added into a translation glossary and/or added to preprocessing for expansion (such as with acronyms). This is accomplished through a combination of NLP, statistical analysis, dictionary lookups, and human translation. Since these different applications all support automotive processing, we have found that a significant number of terms are common across many of these domains. This allows us to use existing glossaries to help build new glossaries. The use of an automotive ontology from the GSPAS system also allows us to identify many of the terms and to consistently handle acronyms and abbreviations.

Application Use and Payoff

The web-based translation system was deployed in early 2010; by the end of that year we were averaging about 2500 translations per day. A translation request is defined as a single transaction that is sent to our system. Typically, this is a sentence or two, but it may also be a longer piece of text or a complete file in PDF, text, Microsoft Word, or other formats. In July of 2012 we sent out a survey to our users to determine their satisfaction with the system and to get feedback in regard to the translation quality. We interrogated the web server log to capture the corporate IDs of those who had used the translation system and sent each of them an automated email with a link to our web-based sur-

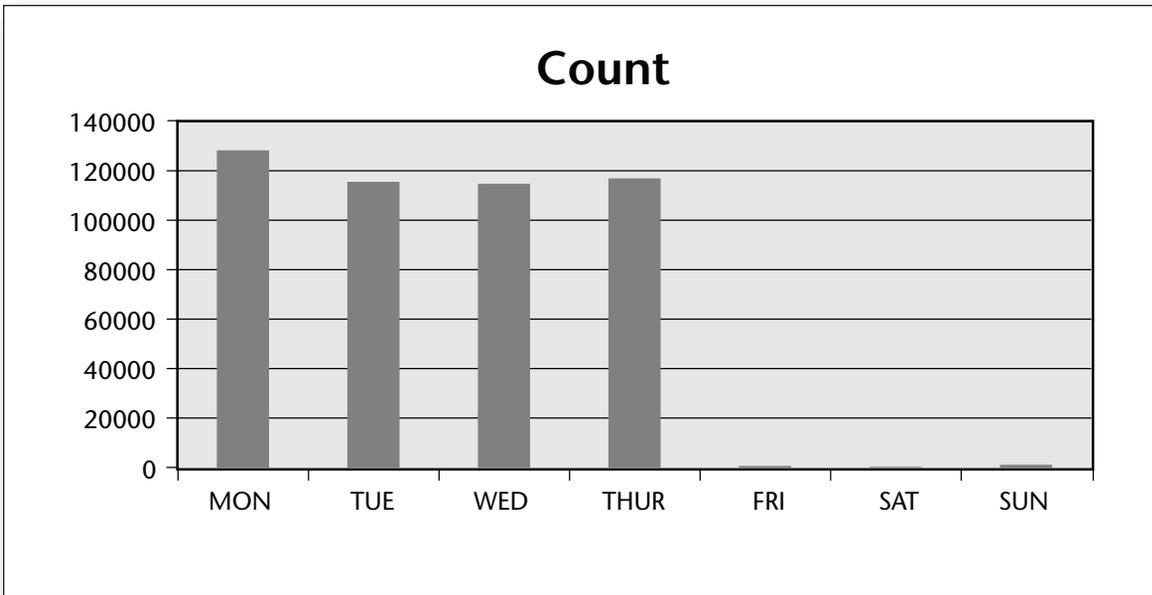


Figure 5. Translation Usage by Day for a Typical Week.

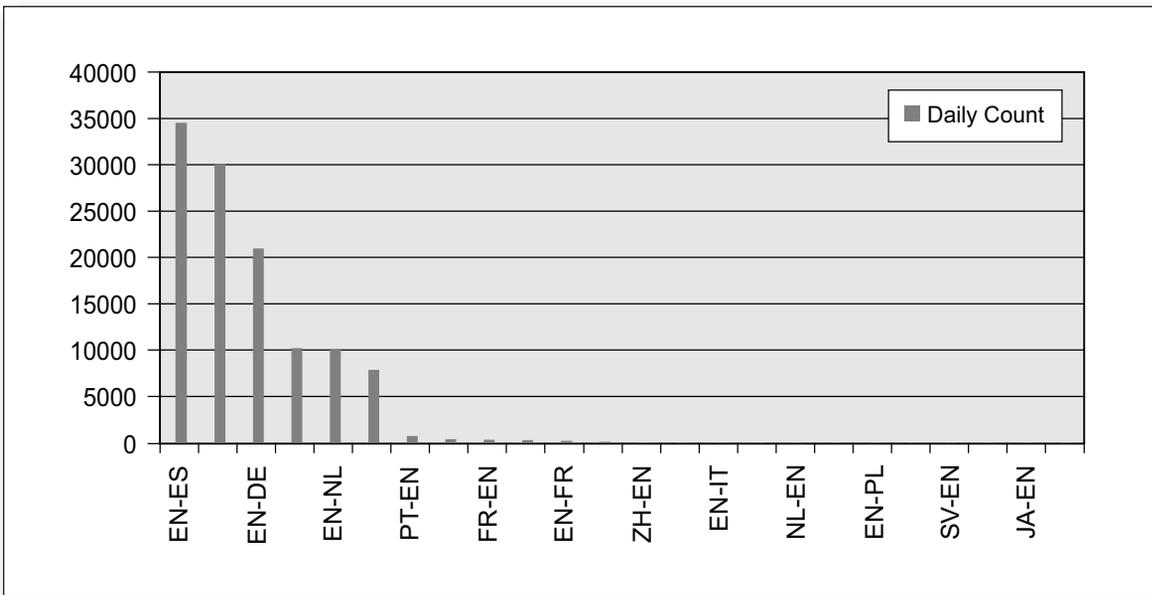


Figure 6. Translation Usage by Language.

Most translations that are sent to the system are a single sentence. We emphasize to the users that complete sentences will translate much more accurately than single words or phrases. In order to translate a single word correctly, the users are encouraged to use that word in the context of a simple sentence. The system also supports the translation of document files, but these are a very small percentage of the overall system usage. Figure 5 shows our translation throughput for a typical week and Figure 6 shows the breakdown of our usage by language. The figure shows that most of our usage is concentrated in a small subset of lan-

guages (Spanish, German, Portuguese, Russian, and Chinese).

The growth in support requests has also shown us that the need for a secure, customizable translation system is an important issue.

Application Development and Deployment

Application development is conducted by a group that is a hybrid between our innovation projects and production support projects. NLP and AI rules

are developed by experts in AI and linguistics and tested in an innovation laboratory environment utilizing real-world data from various groups within the corporation. This lab is firewall protected from the rest of the Ford intranet so that we may test new technologies and deal with confidential data without worry of exposure or impact with existing production systems. Once laboratory innovations are adopted, we hand off developed code/pseudocode and rules to production support personnel who incorporate these changes in a development environment, test for impact on existing code with regression test cases, and then ready for deployment. Deployment consists of bundling glossaries, ontologies, code, and rules and then moving to quality control for testing with production-like data by a separate business-driven group. Finally, approved changes are moved from the quality control systems to production usually during weekend downtime windows (small due to the global nature of the application).

Maintenance

Maintenance on the system is an ongoing process. Our major business requirement for translation is to improve the quality of the translation and to add additional languages into the system. We currently support 14 non-English languages (Chinese, Danish, Dutch, Finnish, French, German, Greek, Italian, Norwegian, Polish, Portuguese, Russian, Spanish, and Swedish) and are planning to introduce additional languages in the near future. The introduction of available commercial languages is a simple process, but the customization for each language requires significant work in order to develop translation glossaries for technical terminology. The glossaries must also be updated as new terminology is always being created and needs to be added into the translation system. Another major issue is the requirement for languages that are not commercially available yet. Many languages, such as Romanian and Thai, do not have a commercial translation engine available, and these need to be developed, which entails additional time and cost. Statistical approaches are effective for developing translation systems for these new languages but require large amounts of parallel corpus data, which is often not available.

Another aspect of translation maintenance revolves around the many existing translation tools and resources that are already present at Ford. Many organizations throughout the company having been independently doing translation using automated tools, external agencies, human translators, and other means. A lot of this work can be reused and integrated into the machine-translation system, but this requires that we find and identify these resources. As new user groups start working with our translation system, we work with

them to develop new translation glossaries and to integrate existing resources, such as translation memories into the Systran tool. To date, we have incorporated more than 25 translation memories into the system to help improve user satisfaction. For a company the size of Ford, this is a huge undertaking, but it will have a large impact in increasing the accuracy of our system and improving the productivity of our employees.

One frustrating aspect of working with machine translation is the difficulty of getting usable feedback from the user community that can be used to improve the product. Many people will continue to manually perform translations instead of contributing to help the automated translation process. Language translation is still a subjective process and people will frequently disagree on what is a "good translation." There are translation metrics that have been accepted by the machine-translation community, but they are complicated and not very useful to the casual user. Our message to these groups includes the idea that even though manual translation by humans has the highest quality, the amount of data that we need to translate can benefit greatly from machine translation with human postediting and a feedback mechanism that can be used to improve the machine rules.

We have developed some methods to quantify the benefits of building translation glossaries through the use of linguistic analysis. Most of the errors that occur in our translations are due to "not found words" where technical terms do not translate correctly unless they are found in the translation glossary. Therefore, we analyze the source text prior to translation and build a table of terms and phrases that do not have translations and then sort these in order of frequency and usage. We can then determine the accuracy of the translations by comparing this list with the translation glossary and calculating how many terms need to be added to the glossary to achieve a specified performance level. This gives the users an estimate of how much work is needed in order to improve the translation quality to an accuracy threshold. This is shown in table 1.

Conclusions and Future Work

The objectives of this article are twofold: to show how machine-translation technology has improved and can be utilized throughout a large company such as Ford and also to show the process by which advanced technologies can be introduced to users across a worldwide company. The biggest change to machine translation has been the growth of statistical and hybrid translation systems (Koehn 2010). This advance has made it possible to develop translation systems much more

Noun Phrases Sorted by Usage Frequency	Count	Percent of Total	Running Percentage
SPOT	5873	4.1%	4.1%
STOCK	4966	3.5%	7.6%
PART	4182	2.9%	10.6%
FIXTURE	3423	2.4%	13.0%
SPOT-WELD GUN	3113	2.2%	15.2%
HOLE	2478	1.7%	16.9%
SCREW	2357	1.7%	18.6%
GUN	2016	1.4%	20.0%
NUT	1680	1.2%	21.2%
BOLT	1589	1.1%	22.3%
SPOT-WELD-GUN	1470	1.0%	23.3%
PALM BUTTON	1405	1.0%	24.3%
VEHICLE	1311	0.9%	25.2%
CLAMP	1143	0.8%	26.0%
BODY	1040	0.7%	26.7%
CLIP	1025	0.7%	27.5%
HAND-TOOL	997	0.7%	28.2%
SEALER BEAD	991	0.7%	28.9%

Table 1. Example of Noun Phrase Analysis.

quickly than with the previous purely linguistic approaches. This is extremely useful for languages that do not have a large commercial or government market and would not normally have been developed. These types of systems can be developed by “training” the system on examples of translated data.

Other advances have occurred in the development of tools that make it easier to test and update the glossaries and to incorporate other translation resources into the system. Our approach to enterprise deployment was based on a bottom-up almost “social networking” approach; we had no mandate to develop a translation solution for the company. We developed a solution within one area of the company and made it available to the rest of the company. There were no processes in place to help deploy the technology across a company the size of Ford; instead, we relied on the many avenues of communication that are open to us now (except mass email marketing) and let the social network communication links that exist spread the word. In many cases, the use of our system did not follow any preconceived patterns. One of the early users of our system was a group in South America even though we had no direct connection with them. But many potential users that were looking for a translation system were not aware of our work for months even though they were located in the same building.

In global companies like Ford, language translation has been an integral part of the business for a very long time. Consequently, there are applications, processes, dictionaries, translation memo-

ries, and other resources available. However, these resources are scattered throughout the company and cannot be easily accessed or utilized. One of our goals is to find and integrate these resources into our system so that the entire company can take advantage of the translation work that has already been done.

Most of our work has focused on translation from English to other languages, but we are now faced with “reverse translation” from other languages back to English. The linguistic analysis tools need to be modified to help process non-English source text and deal with different character sets. There is also a need to translate between languages that do not include English (Spanish to French). The more common languages are available, but many of the less-used language pairs are not available. Translation from one language to English and then to another language will usually not produce very good results. The integration of our world is increasing and communication is an integral part of that process so the need for additional translations will continue to increase.

We have also developed a process to allow computer applications to utilize our translations using a “service oriented” approach. Our interface program handles the communication with the calling system and performs the necessary processing to “clean up” the source text, perform linguistic processing, capture metrics, and process the translation. This solution allows for applications to customize their translation processing and to use a consistent approach for all of the languages that are needed.

Enterprise-level processing at large companies is an extremely complex process. We have demonstrated that the use of AI has made it easier to integrate and customize translation technology into the business processes of our large and diverse user community. Different users have vastly varying requirements and expectations that cannot be met using a “one size fits all” approach. Our approach tried to develop a consistent process to handle these requests and to give the users a level of support that they needed. In some cases, we told users that machine translation was not the right option and that human translators were needed. Managing user expectations is critical; machine translation is not an “out of the box” solution. The quality of the translation crucially relies on the amount of feedback that the user community is willing to provide. Progress is incremental and remains a moving target as our underlying human languages continue to change and evolve. A machine-translation system will not usually provide human-level translation accuracy, but our experience has shown that it has become a technology that our users need and depend on for global business operation.

Acknowledgements

We would like to thank the anonymous reviewers from the Innovative Applications of AI program committee for their comments and suggestions for this article. In addition, we want to thank Markus Fromherz and Hector Munoz-Avila for including this article in *AI Magazine* and for their assistance and many helpful comments that helped to improve this article. Finally, we want to thank our many colleagues at Ford Motor Company who have contributed in making this project a success.

Note

1. www.translate.ford.com

References

- Beregovaya, O., and Yanishevsky, A. 2010. PROMT at PayPal: Enterprise-Scale MT Deployment for Financial Industry Content. In *Proceedings of the Ninth Biennial Conference of the Association for Machine Translation in the Americas*. East Stroudsburg, PA: The Association for Machine Translation in the Americas.
- Depraetere, H., and Vazquez, P. 2010. Machine Translation Engine Selection in the Enterprise. Paper presented at the 2010 European Association for Machine Translation Conference, Sait-Raphaél, France, 27–28 May.
- Dixon, J. 2010. Machine Translation in the Enterprise Environment. In *Proceedings of Ninth Biennial Conference of the Association for Machine Translation in the Americas*. East Stroudsburg, PA: The Association for Machine Translation in the Americas.
- Godden, K. 2000. The Evolution of CASL Controlled Authoring at General Motors. Paper presented at the

Third International Workshop on Controlled Languages (CLAW 2000), Seattle, WA, 29–30 April.

Huijsen, W. 1998. Controlled Language — An Introduction. Paper presented at the Third International Workshop on Controlled Languages (CLAW 1998), Pittsburgh, PA, 21–22 May.

Kardon, A. K. 2005. Application Issues of Industrial Soft Computing Systems. In *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Koehn, P. 2010. *Statistical Machine Translation*. New York: Cambridge University Press.

Matusov, E., and Kopru, S. 2010. AppTek’s APT Machine Translation System for IWSLT 2010. Paper presented at the 7th International Workshop on Spoken Language Translation. Paris, 2–3 December.

Rychtycky, N. 2007. Machine Translation for Manufacturing: A Case Study at Ford Motor Company. *AI Magazine* 28(3): 31–44.

Society of Automotive Engineers. 2002. *J2450 Quality Metric for Language Translation*. Warrendale, PA: SAE International (www.sae.org).

Surcin, S.; Lange E.; and Senellart, J. 2007. Rapid Development of New Language Pairs at SYSTRAN. Paper presented at Machine Translation Summit XI, European Association for Machine Translation, Copenhagen, Denmark. 10–14 September 2007 (www.mt-archive.info/MTS-2007-TOC.htm).

Nestor Rychtycky is a technical expert in artificial intelligence at Ford Motor Company in Dearborn, Michigan. He received his Ph.D. in computer science from Wayne State University in Detroit, Michigan. His research focuses on the application of AI-based systems for vehicle assembly process planning, scheduling, ergonomics, and plant floor applications. Rychtycky is also an adjunct professor at Lawrence Technological University and a member of AAAI and a senior member of IEEE. He chaired the Innovative Applications of Artificial Intelligence (IAAI) conference in 2010 and was cochair for Cybernetics at the IEEE Systems, Man and Cybernetics conference in 2011.

Craig Plesco is a product creation application engineer at Ford Motor Company, specializing in natural language processing. He received his Master’s in linguistics from Wayne State University, in Detroit, Michigan, concentrating in computational psycholinguistics and AI. His work focuses on technological and business futuring of systems that support product creation and related innovation projects including semantic web, machine learning, and knowledge extraction. He is a member of AAAI and the Association for Computational Linguistics (ACL). Recent papers and presentations include his work in language translation (AMTA 2012) and computational linguistic analyses of consumer biases (WSU 2012 Colloquium).