An Effective Model for Insertion of Translation Technologies into US Government Translation Environments

Carol Van Ess-Dykema, Ph.D.[1]
Human Language Technology Research Lead
US Department of Defense
Washington, D.C. United States

Abstract

The [US] National Virtual Translation Center (NVTC) has developed a model for rapid yet incremental insertion of translation technologies into its operational architecture, benefitting from cutting edge technologies, a strong assessment laboratory, key partnerships among international leaders, and day-one collaboration with the NVTC Operations Directorate (its translation customer). This presentation will cover the components of the model for effective technology integration into US government translation environments.

The NVTC Translation Technology Insertion Model

The National Virtual Translation Center (NVTC) has a charter to provide expert, high quality translation for the needs of the United States intelligence, defense, and other federal initiatives. Its role has come to be the translation nexus for tens of thousands of documents, in over 100 languages and more than 20 genres, into and from English. The high quality translations can be used as standalone reports, or as input for follow-on intelligence analysis. The variety of translation requirements at NVTC and throughout the US government challenges its translation technologists to create and implement an architecture ready for the rapid insertion of translation technologies, and a methodology for discerning the best technologies for specific translation requirements.

Such an envisioned translation architecture embodies the orderly introduction of language technology components, human work roles, and the insertion points in the architecture where language technology will be introduced.

The methodology for the Insertion Model takes into account the fact that many types of translation requests will quickly come to be routinely processed through the technologies. The limits on tool availability and maturity for certain languages, however, mean that some translation requests may be accomplished with fewer technology aids. Meanwhile, the capability must grow along several dimensions at once, even as the translation organizations take on larger and larger translation workloads.

---

[1] Dr. Van Ess-Dykema completed her 4-year tenure as Director of Technology at the National Virtual Translation Center during the summer of 2011.

The Insertion Model has five principles:

1. Design and adhere to an advanced architecture which possesses cutting edge technologies, and also the architectural means of interface with future technologies as yet immature or even unidentified.
2. Introduce language translation technology components that will have the largest positive impact. This means targeting language translation technology tools that are mature enough for production use, and which have demonstrated value and relevance in pilot experiments under laboratory conditions.
3. Begin by focusing on automating translation of a few languages end to end.
4. Establish research and development partnerships across communities with translation technology experts.
5. Engage the customer from the beginning through the insertion and beyond.

In this presentation, we will take each principle in the model in turn, focusing upon recent successes and innovations arising from key collaborations with research partners.

Principle 1: Design and Adhere to an Advanced Translation Architecture

The National Virtual Translation Center (NVTC) established the NVTC E-Workflow Translation Technology Online Network (NEWTTON) System, which is unique in the means by which it incorporates translation technologies and feedback mechanisms for best use of technologies and continuous improvement of the translation capabilities (Van Ess-Dykema, 2011).

The NEWTTON system has been endorsed by the Director of National Intelligence through its recognition of NEWTTON as a finalist in the Galileo Awards Program, presented to NVTC in 2010. The Galileo Awards Program is an annual IC-wide competition designed to encourage and recognize innovative workforce ideas that address current challenges and help shape the future of U.S. intelligence.

Figure 1 is a generalized picture of an enhanced architecture, which draws from the experience in developing the NEWTTON system. The translation technology components interact with the relevant components of the overall architecture.
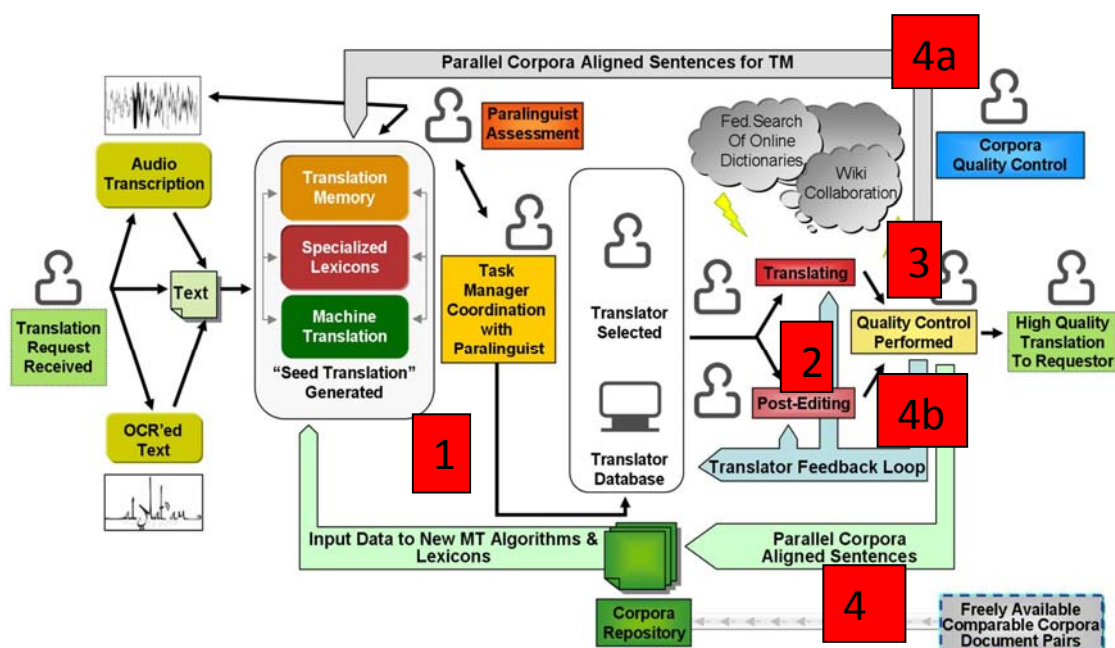
Figure 1. Enhanced translation architecture, portraying the interaction of translation technologies, feedback mechanisms, and the role of the Paralinguist.

The architecture incorporates the configuration of translation technology systems at its front end. The translation enterprise takes in source language in any of a variety of language input forms (text, audio, document images, etc.). The "Paralinguist" possesses general expertise in the application of language technology, as well as specific knowledge of the performance of baseline and customized MT systems, and of the translation memory resources available. The Paralinguist identifies the optimal sequence of integrated special lexicons, translation memory, and machine translation to produce a "seed translation", which is vetted through the Paralinguist and Task Manager, and the appropriate action is chosen for the document (translation or post-editing). The best Translator skill mix for the language direction, topic, and genre is derived from a dynamic Linguist Resources Management System. Translation support includes contemporary collaboration tools and wikis, as well as real-time interaction with the quality control reviewer during the translation process.

Translation technology in support of the translation process is introduced at four points in the workflow (see the corresponding numbers in Figure 1):

Automated "seed translation" generation: project files are logged; image, audio and video may be automatically transcribed; and a "seed" or draft translation is produced using translation memory and machine translation wherever feasible. Relevant terminology is retrieved for each translation, to be provided to Translators as lexical resource packets (Workflow component 1).

Translation / Post-editing process: Translators work from "seed translations" when provided, and have dynamic access to several online resources including a user-driven transcription tool, collaboration with peers and Quality-Control (QC) Reviewer, federated search of dictionary and

terminology resources, translation memory resources, and machine translation (Workflow component 2).

Quality Control: Early QC collaboration enables QC reviewer and Translator to agree on the required style, vocabulary, and register of a document, even as it is being translated (Workflow component 3).

Continuous improvement feedback loops: Following delivery of the complete translation, aligned parallel corpora resources and observations are captured in various systems for future reuse, and process and quality optimization (Workflow component 4). Each feedback subsystem is designed to improve the functionality of a major component of the architecture, immediately at translation time and incrementally as bodies of translation products and judgments are captured and fed back into training models for adaptive algorithms.

Of particular importance to the model are:

- *Feedback to Translation Memory.* Translation memory depends on sentence-level alignments of source language and translated material. To enhance the TM functionality, source documents are aligned with their translations and fed into the TM banks to improve their performance in subsequent translations in the same language pairs, topics, and genres (Workflow component 4a).

- *Feedback to Translator and Translator Database.* This feedback subsystem updates the Linguist Resources Management System. Metrics associated with the Translator's performance and improvements are captured to optimize the use of the best person for each assignment in subsequent engagements (Workflow component 4b).

- *Feedback of Parallel and Comparable Corpora for MT and TM; and Terminology for Specialized Lexicons.* As with TM, MT and specialized lexicons benefit from a feedback of aligned, translated material. Such material can often be fed back into the MT statistical modeling algorithm to improve the translation performance of MT; paired phrases or sentences can populate a TM bank; and specialized terminology can be captured automatically or semi-automatically for Translator reuse (Workflow component 4c).

Each of these subsystems has been the focus of intense research, development, and prototyping by the NVTC and its partners.

Principle 2: Introduce Translation Technology Components that will have the Largest Impact

The model specifies introduction of only the most mature technologies, and a focus on a few of them at a time. The advantage of this approach is the introduction of effective and successful technologies, which fosters user acceptance and willingness for more such insertions.

Selection of the technologies for insertion depends on their current and potential performance for high quality translation needs, which in turn depends on a rigorous set of metrics by which to evaluate them, and a means for performing such evaluations.

Technologies

The components of an end-to-end translation process optimally aided by translation technology need to be shown to interact with each other, with the architecture, and, most importantly, with the human Translators. Thus it is important to have several instances of each such component available for assessment. Figure 2 is a snapshot of the vendor evaluation products and small research prototypes under assessment in the NVTC Translation Technology Assessment Laboratory (Van Ess-Dykema 2011 op. cit.).

MT Systems	Languages
PAHO	Spanish
SYSTRAN	Arabic, Chinese, Farsi, Urdu, Korean
Joshua (University Prototype)	Urdu
SAIC Apptek	Arabic, Chinese, Farsi, Urdu, Korean, Spanish
DARPA/IBM	Arabic
SDL Language Weaver	Indonesian, Arabic, Chinese, Farsi, Urdu, Korean
TM Systems	**Features**
Lingotek	Sentence database, Chinese and Arabic banks; designed for collaboration (US system)
MultiTrans	NLP-assisted document database with sentence links
SDL Trados	Sentence-level database with expert interface
Fluency	30 languages, powerful interface (US system)
OCR Systems	**Features**
NovoDynamics	Recognizes most Arabic-script languages; rapid large volume processing
ABBY Fine Reader	186 languages; interoperates with downstream operations
ReadIris	120 languages; not optimal for batch processing
Adobe Acrobat	Most common languages (not Arabic); widely available and accessible software
Sentence Aligners	**Features**
Terminotix Align Factory	Aligns text while maintaining TM bank attributes
Speech Recognition	**Features**
DARPA/IBM TALES speech-to-text system	Designed for automatic transcription of Arabic news broadcasts

Figure 2.   Snapshot of translation technologies under assessment at NVTC Laboratory.

Metrics

The use of metrics for evaluation is not as straightforward as might be imagined. Many existing metrics for technologies were designed for developers, or for non-Translator users. Others do not cover the variety of conditions (language, document type, etc.) that Translators encounter. Some technologies do not have rigorous metrics at all. Thus a central focus at NVTC has been to identify

the metrics which are the most predictive for the high-quality translation environment, and to develop new metrics when there are none that suit specific needs. To illustrate, we discuss two recent examples of metric development.

TM Technology Metrics Development. As we reported at the Translating and the Computer (31) Conference, (Van Ess-Dykema et al. 2009), a pilot experiment using professional Translators determined which evaluation metrics for TM trended toward scalability to a larger experiment. Such metrics as translation minutes per sentence (with and without using TM), user opinions on ease of use, and correction rates appeared promising and were built into a much larger experiment. In 2010, NVTC conducted its TM Metrics Development Experiment, to determine not only the overall value of TM for high quality translation, but also to develop a reliable and reusable metric that any organization can use to evaluate particular TM systems. The highlights of the large experiment, carried out in partnership with the US Naval Research Laboratory (NRL) and the National Institute of Standards and Technology (NIST) are:

- Translators perform faster and better using TM, compared to the same translation environment without TM; and
- The best metric for reliable, predictive evaluation of TM systems consists of quality control judgments of finished translations.

Paralinguist Decision Factors Development . Key to the success of the Paralinguist role is the ability to determine, nearly instantly, the usefulness of a "seed translation" for post-editing use. The activity is challenging, because of the broad scope of needs (more than 100 languages, 20 genres, into and from English) and requirements for high quality, rapid turnaround translation. The workflow includes a special role of Paralinguist, whose task is to select the best combination of systems to process a source document based on language, genre, and source media of the document. After selecting the optimal human language technology (HLT) tools, the Paralinguist runs the source document through these systems that can include automated speech recognition, OCR, MT, and TM. These tools yield output documents that can be used as "seed translations" for the Translator/Post-editor. The Paralinguist assembles the effective outputs into a packet of translation resources to be passed on to the Translator or Post-editor. US government translation organizations typically translate among many languages, and thus the Paralinguist cannot always judge the usefulness of a "seed translation" generated by MT. For this reason, NVTC has developed a number of Paralinguist decision factors which will automatically inform the Paralinguist about the usefulness of each "seed translation."

NVTC performed an experiment to determine which properties of MT might be a diagnostic predictor of the potential for the "seed translation" to be post-edited. This experiment proceeded from two hypotheses: Translators will perform better post-editing a "seed translation" than translating from scratch; and that Paralinguist decision factors can be determined to assess whether a given MT output is suitable for human post-editing. The experiment found that Translators perform generally better by post-editing MT output than by translating from scratch (both in time and quality), and that there are characteristics of documents (whether in source form or "seed translation" form) which are strong predictors of post-editing suitability. Among these are:

- The length of the source document;
- The average word length in the source document;
- The format of the source document;
- the ratio of source-to-target sentence length in a "seed translation" compared to the ratio in human translation in the same language pair;
- Number and percentage of not-translated words in a "seed translation."

The NVTC Translation Technology Assessment Laboratory

The NVTC Translation Technology Assessment Laboratory became operational in 2010, and has since been the site of a large number of experiments, evaluations, and user assessments. At any given moment, the Lab has several MT systems, TM systems, and OCR systems, with the appropriate interfaces among them to enable a timely assessment of the value of any particular combination of the technologies for any specific translation purpose. Researchers and NVTC professional Translators use the laboratory to assess technologies and processes for enhanced translation.

Activities in the Lab are guided by concise, actionable test plans, and a clear definition of the scope of the activity (expert/user assessment, trend study, or statistically valid full experiment). Typically, assessments are performed by NVTC's own operational Translation staff, which has helped define evaluation parameters and insertion requirements for the architecture.

Principle 3: Prototype Language Capability End-to-End

The NEWTTON architecture and the NVTC Laboratory are both configured to reflect the fact that the implementation of the architecture is incremental. The translation technologies must interface with other components of the architecture, such as the document flow through the human work roles. Prototyping for this complex capability requires forethought and planning.

The individual capability of a particular translation product is only as useful as its ability to be integrated into the sequential process. Thus it is important to determine which technologies may be loosely integrated (e.g., OCR), and which must be more tightly integrated (TM), and which are variable (terminology management).

At the same time, it is important to decide on the best language pairs for prototyping, since these will also be the first pairs to be implemented in the operational architecture. There are three considerations:

Current and projected customer needs. The individual customers know their immediate need, but viewed from the needs of the entire US government, the languages are much harder to identify, particularly given the rapid changes in regional events with international implications.

Availability of large training corpora. Prototyping also requires a sufficiently large volume of parallel corpora. Even though NVTC often translates low resource languages, the prototypes need to be supported by the requisite amount of test data for each technology in the architecture and the feedback loops.

Availability of a sufficient body of technologies to prototype end-to-end strategies.   As with the need for high volume language data, there needs to be a suitable number of instances of each technology product type, in order to determine both the best product and the best way to integrate it into the architecture (See Figure 2).

Principle 4: Establish Research Partnerships

NVTC has benefitted from opportunities to team with US and international organizations – from research institutions to government entities to commercial developers – to help research and prototype portions of the architecture, particularly the Paralinguist decision factors, the feedback mechanisms, and human factors interfaces.

DARPA

Perhaps foremost among these is the strategic relationship with the US Defense Advanced Research Projects Agency (DARPA) based in common goals and in common understanding of the potential for translation technologies.   Since its inception in 2005, the DARPA Global Autonomous Language Exploitation (GALE) program has endeavored to combine speech and text MT to provide actionable information to military command and personnel in a timely fashion. NVTC is engaged in several activities under the GALE aegis:

- *SDL Language Weaver*. Synergistic with the in-house Paralinguist Decision Factors development described above, SDL Language Weaver is working to provide automatic translation quality prediction for paralinguistic decision support for NVTC. A critical factor in the Paralinguist tool suite is the value of an MT output, particularly in the common event that the Paralinguist does not know the target language (Van Ess-Dykema and Reeder 2010). For this capability, it is developing a binary decision framework for the Paralinguist, based on its existing five-point TrustScore, originally designed for determining publication quality of raw output.

- *IBM Research.*   The feedback subsystems shown in Figure 1 are intended to enable a rapid improvement of MT "seed translations" by using NVTC human translations as data to remodel the MT engines. IBM Research is prototyping this key component of the NVTC NEWTTON architecture, for Arabic and Urdu.

  Under the DARPA partnership, IBM Research has also provided the NVTC Laboratory with its Arabic-English statistical MT system, and its Translingual Automatic Language Exploitation System (TALES), for automatic speech and video recognition in Arabic.

The DARPA Broad Operational Language Translation (BOLT) program will begin in October 2011 as a complement to GALE, providing translation in support of Defense Department needs for a wide area of language contexts, from single phrase translation to the translation of large data sets of voice, video and print.

International Workshop on Translation Memory Technology

A milestone event arose out of NVTC's TM activities: an international workshop on TM held in the Washington, D.C. area in May 2011. International participants came from the National Research Council Canada, the Pan American Health Organization, the International Monetary Fund, and the World Bank. US Government organizations, academia, and professional translation companies joined the international representatives in an animated series of discussions on the use and future of TM and its integration with machine translation.

Julia Aymerich from Pan American Health Organization (PAHO), Washington, D.C. ably delivered the first keynote address, explicating the history of the PAHO MT system and its recent exploration into TM integration. Salim Roukos, Chief Technology Officer of Translation Technologies at IBM Watson Research Center, told the IBM story of how IBM integrated translation memory into its processes for translating its own technical manuals.

The NVTC and Naval Research Laboratory lead scientists presented details of the completed TM Metrics Development experiment, with the conclusions noted above. The energized participants of the workshop engaged in dialogue about the experiment methodology and findings, and their potential for using the new TM metric in their own organizations.

Pierre Isabelle, National Research Council Canada, and Daniel Marcu of SDL Language Weaver led a research panel on issues of deep integration of TM and MT to capture the strengths of both technologies, currently and in the future.

Senior managers from the Office of the Under Secretary of Defense for Intelligence, the US Department of State, NVTC, and other agencies shared their perspectives on the use of translation technologies in the US government.

Partnerships with Strategic Research Laboratories

NVTC architecture development is supported by substantial research prototyping and assessment. These technical inquiries on behalf of NVTC require the expertise of strategic laboratories in the translation technology field.

[US] Army Research Laboratory (ARL). Scientists from ARL were able to determine the range of effectiveness of parallel corpus feedback, in conditions of very low resource language material and few translators. Their work demonstrated that MT models built from chapter-length increments of a human translation of a military manual improved the machine translation of the subsequent chapters of the manual.

[US] Naval Research Laboratory (NRL). Beyond its work with the TM metrics development, the NRL team has taken the lead in investigating the potential of TM-MT integration for the translation architecture. Their approach involves investigating alternatives, from sequential integration of commercial TM and MT systems, to the sentence-by-sentence interaction of the TM and MT engines in a single system. The NRL findings will have a significant impact on the way US government translation enterprises plan for their translation technology implementation.

National Research Council Canada (NRC). NVTC has engaged the Interactive Language Technologies Group at the NRC Canada to extend their proven approach to document categorization for information retrieval to document genre identification for translation technology (Simard et al., 2009). NVTC researchers theorized that TM is sensitive to genre (journal article, user manual, email) in the same way that statistical MT is sensitive to topic (medical, military, finance). The challenge for NRC was to automatically determine the genre of a source document in a very large corpus of documents, across languages. Among NRC's findings is the rather surprising result that the statistical distribution of function words (articles, prepositions, and conjunctions) and punctuation in a document is nearly as good a predictor of genre as human judgment. NRC continues to work to predict the usefulness of MT or TM for translating a particular document based on its genre.

University of Maryland Center for Advanced Study of Language (CASL).[2] The CASL Language Use Research Team is providing human factors expertise to several challenges in government translation. Chief among these is capturing the assessment of the Translator in determining the best products in processes in automating translation. Product assessments that yield statistically significant results require large numbers of human experts, which is costly and time-consuming in a government context. CASL is devising methods for equipping the professional Translator to make product assessments that yield highly reproducible, predictive assessments, alleviating the need for large number of Translators in a controlled experiment. CASL is using best practices from the field of subject matter expert assessment, along with plan-based scenarios for optimizing the expert Translator's exposure to new technology.

Principle 5: Engage the Customer from the Beginning

The Operations Directorate of NVTC is constantly performing translations whose timeliness and accuracy are critical for the United States defense, intelligence, and safety of the homeland. It is crucial, then, that the introduction of new technology occur with no downtime, and little or no ramp-up time. At NVTC and at any translation organization, all of the innovations planned and implemented need to be performed as collaboration between technology and operations, ensuring a continuity of workflow function as cutting edge technology is introduced.

The functions of Operations and Technology intersect at several critical points, including:

- The interaction with the systems integration involved with implementing the architecture;
- Definition and establishment of the new work roles into daily practices; and
- Familiarizing Translators and Managers with the candidate technologies and the improvements they bring to the translation process.

Extending the Model for Technology Insertion into other US Government Environments

---

[2] CASL is a US Department of Defense University Affiliated Research Center (UARC), located at The University of Maryland.

NVTC has adopted the NEWTTON architecture and has already implemented features associated with the new human work roles and the feedback mechanisms. It has identified the best suite of OCR systems for its language needs. It developed the metric for TM technology assessment, and has a plan for using it to select one or more TM systems. Processes are in place for aligning parallel corpora and for MT as an adjunct to OCR.

There are many translation organizations within the US government. NVTC is sharing its progress and successes with these organizations who may decide to implement architectures similar to NEWTTON and which are responsive to their individual missions. The principles described here are extensible to meet both the general needs that translation organizations share, and also the unique aspects of their individual responsibilities.

Conclusion

Implementation of a new architecture for a high-throughput, high quality translation provider requires strict adherence to a set of principles, a commitment to provide the best technologies to the Translator, reliance upon strong partnerships in the research community, and a close collaboration with the translation customer.

Acknowledgement

I would like to thank Dr. John S. White, MITRE Corporation technical support, for his substantial contributions to this paper. I would also like to acknowledge his contributions to the experimentation that took place in the NVTC Translation Technology Assessment Laboratory and to the development of NVTC's partnerships with the strategic research laboratories.

References

Simard, Michel, and Pierre Isabelle. 2009. Phrase-based Machine Translation in a Computer-assisted Translation Environment. Proceedings of the 2009 Machine Translation Summit. Ottawa: International Machine Translation Association.

Van Ess-Dykema, C., Perzanowski, D., White., J., and Converse, S. 2009. "Exploring Translation Memory for Extensibility across Genres: Implications for Usage and Metrics." *Translating and the Computer 31. Proceedings of the Conference.* November 19-20, 2009 London, UK.

Van Ess-Dykema, C., Reeder. 2010. "Paralinguist Assessment Decision Factors for Multi-Engine Machine Translation Output: Post-edit or translate." *Association for Machine Translation in the Americas (AMTA-2010)*, *Government Session.* November 1, 2010. Denver, Colorado.

Van Ess-Dykema, Carol. 2011. Improved Workflow and Technology Metrics at the US National Virtual Translation Center. Joint International Annual Meeting on Computer-Assisted Translation and Terminology (JIAMCATT 2011). Turin, Italy: United Nations.