

Week Twelve Response

Machine Translation: not now, but someday

At present, Department of Homeland Security (DHS) funding of a Machine Translation (MT) project is premature. The volume of translation work that needs to be processed by the DHS is immense, and any approach that might speed up the handling of this material would be of great importance. Every day, a literally countless number of communications are intercepted which require translation, possibly decryption, interpretation, and classification based on their content. The present day state of translation protocol and resources makes true MT such a distant goal that its importance is vastly outweighed by the current demands on the Department.

In its effort to prevent terrorism and to protect the nation's interest, the DHS must filter through a vast number of communications intercepts in a variety of languages. Additionally, these intercepts must be parsed for key words or phrases that have significance to security matters. Currently this task is accomplished through a combination of methods, in-house experts, assistance from other agencies (such as the NSA), and outsourcing to private contractors. This mix of approaches results in a logistical nightmare, as well as opening the door to security leaks. If the DHS were to have a Machine Translator, it would be possible to handle virtually all of our translation needs in-house, and we would only need to go outside of our department for specialist consultations.

This is a dream in the intelligence community that has been around since the middle of world war two when people like Alan Turing pointed out that if computers could crack codes, they should be able to translate from a foreign language. The view then, which is still held by many today, was that a foreign language is just another code. This dream of an "ultimate spy computer" that could replace human expert translators with higher accuracy, speed, efficiency and consistency has been in the minds of the scientist and intelligence professionals for roughly sixty years. Similarly, since the appearance of personal computers on the market, translators themselves have looked for ways to make this powerful tool work for them. The best that they have been able to do to date has been the increased availability of simple resources (such as dictionaries) to the general community. Today few tools exist for translators that truly take advantage of the computer's abilities.

One of the key problems is that languages are fundamentally not like codes. There are certain subtleties that defy simple translation and representation in foreign languages. Another problem is the used of simple digitizing methods. Any intercepted communication can be more easily manipulated if it is first digitized; this also increases the ease of communication and collaboration on an important intercept. The methods for this digitization are riddled with faults: optical character recognition and speech recognition software are less than reliable in English never mind the numerous dialects and transcriptions of Arabic and other languages based on character sets other than the Roman alphabet.

The golden standard in Machine Translation is Fully Automatic High Quality Translation (FAHQT), but it is experts' consensus that such a machine is currently not possible. The best reputed translators are 98% accurate (granted these claims are suspect, and the metric for translator accuracy is by no means an industry standard), but even at that seemingly high level, a sizeable number of errors sneak through in each document. Over the course of a single day at the DHS, this number would mount up to an

Week Twelve Response

astronomical figure. Furthermore, the kinds of errors that do slip through would be far more egregious because the computer by design does not make “simple mistakes” much as a human counterpart might do. This begs the question of how much of a *Safety Critical* system is DHS translation. Do lives depend on its accuracy? Absolutely, but how much of a margin of error are we willing to tolerate in the field, are we ready to take a less than perfected weapon out on to the battle field in the war on terror?

All these facts point towards the perhaps fatal consequences of depending on a method of Machine Translation based on current technology. This development is too far off to be handled by the DHS with its pressing agenda. A compromise is however in order, incorporating many of the components that will be necessary for a future MT system instead to assist human translators. Computers are still decades away from truly being able to translate on a functional level, but the base technologies are finally starting to blossom. If these are made available to a centrally organized translating service within the government, great strides in national security can be made. In the meantime, the problems of system expertise must also be addressed. Any system that can be an effective translator for the DHS, the military, the NSA or similar groups must be able to use expert knowledge in the fields of medical, scientific, engineering, cultural, religious, geographic and historical information in order sensibly. There is hope for future MT, but its development should be left to academics and industry for the meantime.

Week Twelve Response

Works Cited:

Erard, Michael. "Translation in the Age of Terror." *Technology Review* (March 2004): 54-60.

Hutchins, W. John. *Machine Translation: Past, Present, Future* (Chichester, UK: Ellis Horwood, 1986), chap. 2, "Precursors and pioneers"; chap. 8, "Expectations and criticisms"; chap. 19, "Present developments and some future prospects."

Routledge, John Newton. *Computers in Translation: A Practical Appraisal*, London, 1992. <http://language.home.sprynet.com/lingdex/limtran1.htm>

Weizenbaum, Joseph. "The Computer and Natural Language." Chap. 7 in *Computer Power and Human Reason* (New York: Freeman, 1976), pp. 182-201.