USAWC STRATEGY RESEARCH PROJECT

# Machine Translation: A Key to Information Supremacy and Knowledge-Based Operations

by

James Koh
Department of Army

Professor Kevin Cogan, COL (ret.) USA
Project Advisor



U.S. Army War College
CARLISLE BARRACKS, PENNSYLVANIA 17013

# REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188



| 1. REPORT DATE (DD-MM-YYYY)<br>07-04-2003 | 2. REPORT TYPE | 3. DATES COVERED (FROM - TO)<br>xx-xx-2002 to xx-xx-2003 |
|---|---|---|

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>Machine Translation: A Key to Information Supremacy and Knowledge-Based Operations<br>Unclassified | 5a. CONTRACT NUMBER<br>5b. GRANT NUMBER<br>5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S)<br>Triplett, Mark ; Author | 5d. PROJECT NUMBER<br>5e. TASK NUMBER<br>5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME AND ADDRESS<br>U.S. Army War College<br>Carlisle Barracks<br>Carlisle, PA17013-5050 | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS<br>, | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
APUBLIC RELEASE
,

13. SUPPLEMENTARY NOTES

14. ABSTRACT
See attached file.

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| | | | Same as Report (SAR) | 39 | Rife, Dave<br>RifeD@awc.carlisle.army.mil |
| a. REPORT<br>Unclassified | b. ABSTRACT<br>Unclassified | c. THIS PAGE<br>Unclassified | | | 19b. TELEPHONE NUMBER<br>International Area Code<br>Area Code Telephone Number<br>DSN |

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39.18

# ABSTRACT

AUTHOR: James Koh

TITLE: Machine Translation: A Key to Information Supremacy and Knowledge-Based Operations

FORMAT: Strategy Research Project

DATE: 07 April 2003 PAGES: 39 CLASSIFICATION: Unclassified

Informational Globalization unleashed by the recent advent of information technology has brought the world closer than ever by placing the world on a single information grid. Ironically, abundance of data makes information a much more serious and important commodity. This is because access to data is no longer limited to those few well-endowed nations, but others who did not previously have such privilege. This presents a new challenge and opportunity. The value of information is based not solely on its content and accuracy, but also on its speed of acquisition. Acquiring relevant and accurate information from data before others often decides a victor. From this perspective, information globalization is about information competition that turns data into information and knowledge.

Now, more than 60 percent of data on the Internet is from foreign origins, often in their own languages. That percentage is rapidly increasing. This puts those who are not proficient in foreign languages a great disadvantage in terms of data understanding and acquisition speed. How will the U.S. cope with this challenge and achieve information supremacy now and in the future? What are the current U.S. foreign language capabilities and what are the requirements? Do current capabilities fulfill the requirement? If not, what are the potential alternatives? Can 21st Century technology be a solution? This paper addresses these questions. It explores whether Machine Translation technology can provide a key to the Information Supremacy and Knowledge-based Operations for the Nation.

## TABLE OF CONTENTS

## PREFACE

First, I would like to thank my project advisor, Prof. Kevin Cogan, for his insightful guidance and advice. I would also like to thank two of my classmates at the AWC, Col (P) David Heininger, USAF, and LTC William Morris, USA, for being second readers and their comments. In addition, I would like to recognize CDR Robert Noelsch from my previous office, Combatant Commander Interoperability Program Office (CIPO), for providing me with wealth of information.

## LIST OF ILLUSTRATIONS

# MACHINE TRANSLATION: A KEY TO INFORMATION SUPREMACY AND KNOWLEDGE-BASED OPERATIONS

"Lingua Est Potentia"

(Language is Power)

- Koh

## INTRODUCTION

### POWER OF LANGUAGE

> The whole world had one language and a common speech. They said to each other, 'Come, let us make bricks and bake them thoroughly.' They used brick instead of stone, and tar for mortar. Then they said, 'Come, let us build ourselves a city, with a tower we may make a name for ourselves and not be scattered over the face of the whole earth.' But the Lord came down to see the city and the tower that men were building. The Lord said, 'If as one people speaking the same language they have begun to do this, then nothing they plan to do will be impossible for them. Come, let's go down and confuse their language so they will not understand each other.' So the Lord scattered them from there over all the earth.[1]

The story of the Tower of Babel from Genesis, a familiar one to many, though notwithstanding any regards of its accountability and religious faith, epitomizes the power of language as an instrument to knowledge. Language enables and fuels human activities. Advancement of mankind is a product of accumulation of human knowledge over time, and would not have been possible and would have perished without language to communicate and record. Reinventing the same wheel would occur in every generation without language.

The power of language validates itself in this age of globalization as it has been unleashed by information technology (IT). Information becomes more important than ever. Speed and efficiency of acquiring knowledge (*knowledging*) become even more important. Webster's dictionary defines *knowledge* as the sum of what has been understood, discovered, or learned.[2] In this sense, data is not information or knowledge. It must be translated and transformed to be relevant. Language understanding plays a vital role in this process. For instance, information presented in Farsi would mean little to those who do not understand the language: no *knowledging* process occurred.[3]

This paper's focus is on foreign language capability and its implication for U.S. national and military security strategy. This thesis follows on the theme by asking three primary questions: 1) is there a foreign language capabilities requirement in the U.S. national and military security strategy? 2) is the requirement currently being met by the capabilities? and 3) can technology be an answer?

## FOREIGN LANGUAGE AND GLOBALIZATION

No single common language exists with which the whole world can intuitively communicate, exchange and share, and understand individual thoughts. Esperanto, "*one to hope,*" was an attempt at bringing the world under a common and neutral language, designed to facilitate communication without any boundary of eco-politics.[4] At the time of introduction of Esperanto in the late 19th century, internationalism was in fashion, sweeping the western countries. Technologies of that era brought differing regions of the globe closer to one another. However, despite this favorable climate for success, Esperanto was perceived as only an idea and did not really get off the ground. Learning this "common" language was difficult, complex and unnatural. In addition, it was overlooked that language parallels and reflects the real world. Dominant powers did not see a compelling need to acquire a new language skill to *communicate* with third-world countries. English, Dutch and French were the standards and continued to prevail. The "*one to hope"* had folded the hope.

Today, there is a different attempt at a common language. It is called Globalization. This time, the idea of *one to hope* may come about as capabilities of supporting information technology of today far exceeds that of 19th century. Many describe their understanding of what globalization may be in various terms such as "inexorable integration of market, nation-states, and technologies which enable individuals and nation-states to reach around the world farther, faster, deeper and cheaper than ever before;"[5] "the compression of the world and the intensification of consciousness of the world as a whole;"[6] and "the historical transformation constituted by the sum of particular forms and instances of . . . [m]aking global by the active dissemination of practices, values, technology and other human products throughout the globe."[7] In short, globalization can be summarized as individuals and nation-states reaching out and touching information. From a technological point of view, information globalization is about placing the world on a common information grid in which the classical meaning of information divergence and convergence means little.

Information everywhere is waiting to be mined, although it may not be written in English. For instance, one website written in one language may be accessed by many people from all around the world, as they translate the content into their own native language.[8] Non-English language on the World Wide Web (WWW) is now estimated as approaching 60% of the total, and its total percentage is growing even larger. There is no doubt readers prefer to have text in their own language, no matter how flawed and error-ridden it may be, rather than to struggle to understand a foreign language text. Also significant will be the growth of multilingual access to information sources. Increasingly, the expectation of users is that on-line databases should be searchable in their own language, that the information should be translated and summarized into their own language.[9] This makes foreign language capabilities an imperative enabler in information globalization. Paradoxically, information globalization that expands the rim of information increases competition for information. No longer is the focus on who gets data first, but it is more important who analyzes, process, understands it first – *knowledging process*. An old adage, "knowledge is power" seems truer than ever before, ironically in the era of information technology.

## FOREIGN LANGUAGE AND U.S. SECURITY STRATEGY

How does the U.S. as a nation meet the demand for achieving the national security objectives in the world of rapid globalization? As a globalization process diversifies the world, so does the need for foreign language capabilities. This is mainly attributed to the diversification of information sources and a changing security environment in light of such events like the terrorist attacks of September 11, 2001. A recent report reveals that foreign language capabilities in the U.S. federal government are approaching dangerously low levels potentially impacting the mission of protecting our national security and interests.[10] The ability to communicate with other national security agencies to interdict drug trafficking, monitor terrorist activities, and conduct coalition military operations is vital to securing the national security objectives. Adequate foreign language capabilities are a must to support traditional diplomatic efforts and public diplomacy programs, military and peacekeeping missions, intelligence collection, war on terrorism efforts, and international trade. It is key to successful and effective diplomacy, defense, and intelligence gathering.[11] For that reason, the Department of Defense (DOD) foreign language capability needs for national security are driven by the National Security Strategy and the National Military Strategy.[12] DOD estimates that it alone currently spends up to $250 million annually to meet its foreign language needs.[13]

However, foreign language capabilities are critical not only to the DOD, but also to other Government agencies. The Department of State personnel testified to the congress that the shortfalls in foreign language proficiency have contributed to a lack of diplomatic readiness. As a result, the representation and advocacy of U.S. interests abroad has been less effective. U.S. exports, investments, and jobs have been lost and the fight against international terrorism and drug trafficking has been weakened.[14]

The intelligence community echoes concern on this issue. The primacy of foreign language skills cannot be overstated to the community's core mission. It is critical in all phases of the intelligence cycle from collection to exploitation to analysis and production. Information or input may come in different languages and sources which need to be interpreted and analyzed rapidly. Currently, the intelligence community does not always have the available resources to meet such requirements. With the end of the Cold War and the ensuing movement towards globalization, the threats have become more complex and diversified, which has increased foreign language needs.[15] This complexity and diversification has weakened the U.S. foreign language capabilities to fighting against international terrorism and drug trafficking and resulted in less effective representation of U.S. interest overseas.[16]

The prospect for meeting the needs of the intelligence community on the foreign language capabilities is unfortunately troublesome. The Federal Bureau of Investigation (FBI) may lose more than half of its linguists and international experts through retirement in the next five years. This will leave the FBI with significant shortfalls of personnel needed to investigate international organized crime.[17] About a decade before the horrific September 11, 2001 attack, the World Trade Center was targeted for a terrorist attack by radical followers of an Egyptian sheik. The terrorist group used a code word "Hadduta" for the bombs, which means 'children's bedtime story' in Arabic. Fortunately, the FBI who conducted the surveillance understood the language, deciphered the code, and seized the Islamic radicals. However, and alarmingly, the FBI may not have the same capability for the future if large portions of their language expertise are lost through retirements. Further degradation of foreign language capabilities presents serious implications from a national security standpoint.[18]

DOD shares the same concern. The 2002 National Military Strategy assumes superior information and knowledge of its operations as a major tenet to the full spectrum capabilities. It also views small-scale contingencies (SSC) and peacekeeping operations (PKO) in various parts of the world as encompassing the predominant forms of future U.S. military operations. U.S. forces will operate with coalition forces and foreign civil organizations in environments in which different languages, cultures, and religions dominate. The ability to communicate clearly in such operational environments with allied and coalition forces and with current and potential adversaries is imperative for mission success. Foreign language skills are required to conduct effective interactions with allied, coalition, and host-nation forces while facilitating intelligence, civil affairs, psychological operations, and military training.[19] Unfortunately, the U.S., although nationally pluralistic, does not have the range of native or learned linguists in its military forces to meet such linguistic requirements. At any one time, the total U.S. military needs are estimated to be 30,000 civil employees, contract translators, and interpreters dealing with over 80 different languages. Combatant Commanders have reported significant shortfalls.[20] For example, on-going peacekeeping operations in the Balkans generated significant language requirements and revealed a significant shortage of organic linguists in the services. Just for the Balkan operations alone, the DOD hired more than 900 linguists on contract to meet the requirements. Defense contractors who needed to provide linguists to DOD experienced difficulty in recruiting qualified personnel to the positions while use of non-U.S. Government personnel raised security concerns.[21]

At the component service level, the Army has considered five languages critical: Arabic, Korean, Mandarin Chinese, Persian Farsi, and Russian. The Army had authorizations for 329 military translator and interpreter positions for these five languages in fiscal year 2001 but only filled 183 of them, leaving a shortfall of 146. In addition to its needs for translators and interpreters, the Army also has a need for filling staff positions with applied language skills. Two key job series involve military intelligence – cryptologic linguists and human intelligence collectors. The Army had a shortfall of cryptologic linguists in two of the five languages deemed most critical – Korean and Mandarin Chinese. It also had a shortfall of human intelligence collectors in all five foreign languages. As a result, the Army has noted that a lack of linguists is affecting its ability to conduct current, and anticipated human and signal intelligence missions. Consequently, the Army said that it does not have the linguistic capability to support two concurrent major theaters of war.[22]

Thus far, foreign language capabilities requirements and deficiencies, and its implication to the national and military security strategy were discussed. The federal agencies and departments are searching for ways to improve the situation. Their main approach seems to gravitate towards the traditional approach of instruction: the Defense Language Institute and the State Department's National Foreign Affairs Training Center. Both government operated institutions offer the best language training available. However, acquiring foreign language skill is more an art than a science. It simply takes time to learn a language. Current training programs most likely would not produce the number of linguists with sufficient skills in the desired timeline. Is there an alternative? Would technology offer utilities to improve the situation?

## MACHINE TRANSLATION

### INTRODUCTION

In simplest terms, Machine Translation (MT) is having computers translate texts from one natural language to another, for instance, from Russian to English or Farsi to Chinese. MT is a part of human language technology with many variants and is a subject researchers, scholars and engineers wrestle with as to which MT approach is best. It is understandable since human language activities are difficult to assess, quantify, model, and emulate with few formulae and machines.[23] Machine translations can either be fully automated as is MT or semi-automated, known as Machine Assisted Translation (MAT). The main difference between MT and MAT lies in whether translation is performed with or without human interaction such as pre-editing of the input text to the translation machine or post-editing of the output from the machine.[24] The main argument for needing such distinction has to do with its applications, utilities and user groups. Some argue that the failure to identify different needs and to design systems specifically to meet them has contributed to misconceptions about translation technology and its impact for the professional translator.[25] However, this paper will not make the distinction since a perfect system has yet to arrive, and pre- and post-editing would certainly improve the quality of translation greatly. Additionally, some applications would not even need editing. These include a key word search, data mining, and short and very descriptive control words. Even the most mediocre MT system can outperform those areas with no sign of fatigue.

### HISTORY

The idea for MT dates back to the 1940s as Warren Weaver of the Rockefeller Foundation approached a text written in Russian as if it were written in English having strange symbols and codes, just like cryptology would approach an encoded message. His idea was to build a machine to automatically decode or *translate* the text so that meaning of the text can be extracted.[26] By end of the 1950s, a group of researchers mainly in the U.S., Russia, and Europe followed the idea. They felt that they could develop high-quality MT systems within a few short years, capable of translating scientific and technical documents.[27] To their disappointment, they soon realized how complex and difficult a problem it would turn out to be. In 1966, the National Academy of Science’s Automatic Language Program Advisory Committee (ALPAC) which, had funded many of the MT programs, recommended that funding for MT should be redirected more towards the fundamental question of computational linguistics before any practical translation machine could be built. The MT community was sharply divided by the recommendations. However, the recommendations were adopted and as a result many laboratories cancelled MT projects while some shifted their research focus to long–term research in computational linguistics.[28] By 1973, only three government-funded programs were left in the U.S., and by 1975 there were none.[29] In spite of canceling all funding from MT projects, U.S. governmental agencies continually used early versions of MT systems as the only alternative to human operators for information gathering from the Soviet Union.[30] They simply did not have an alternative to MT systems which were able to process significant information in a short period of time. In particular, the multilingual communities of Canada and Europe emphasized the urgent need for numerous levels of translation production, far beyond the capacity of the professional linguistic community. It was quite clear that some help from computers was a necessity.[31] There was a resurgence of interest in MT in the 1980s, notably in Japan. Promising results were based not only on linguistics, but also on the power of a new generation of computers and engineering minds on approaching MT.[32]

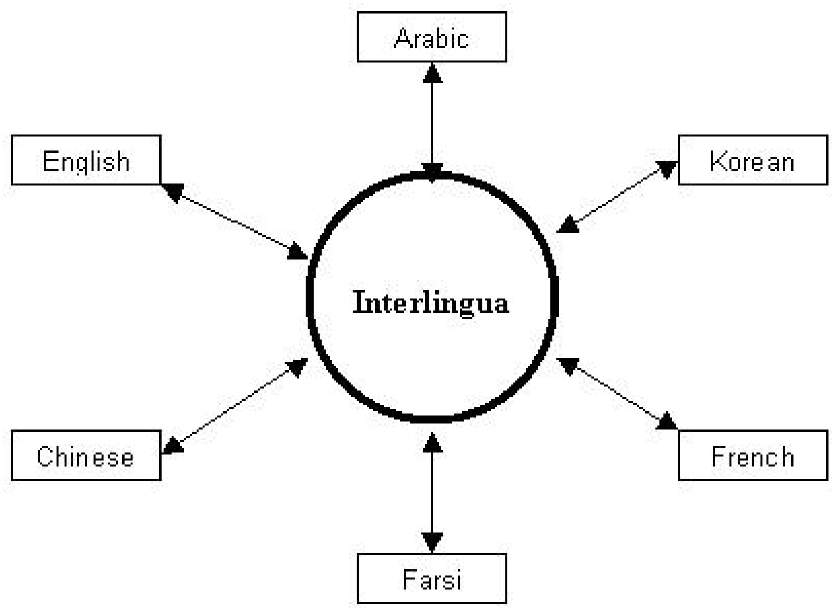
## TECHNOLOGY

Natural Language Processing (NLP) is closely linked with linguistics and depends on many linguistic language theories. The attempt to process natural languages using computers is not as easy as it sounds. In fact, natural language is a very difficult equation for computers to deal with effectively.[33] Since the 1960s, three major approaches to MT have dominated the MT community: Direct, Interlingua, and Transfer.

The Direct approach involves the direct swapping of words and structures from the source to the target language with minimal disambiguation operations. For this reason, the Direct approach can be successful only with similar language pairs which have similar grammatical structures. For dissimilar language pairs, the Direct approach translation can be quite inaccurate because the number of equivalent words and phrases between the languages may be insufficient.[34]

The Interlingua approach is similar to the concept of Esperanto. Interlingua is a conceptual representation of meaning, independent of any language. It presumes that meanings are language independent. For example, different languages describe the word "beautiful" differently, but all mean "beautiful." In other words, Korean word "*Areumdaun"* and English word "beautiful" clearly have a different way to express (representation) the meaning, but both have the same meaning, "*beautiful*." If the representation of the meaning for "beautiful" in any language can be translated into a conceptual representation, then it is called Interlingua.

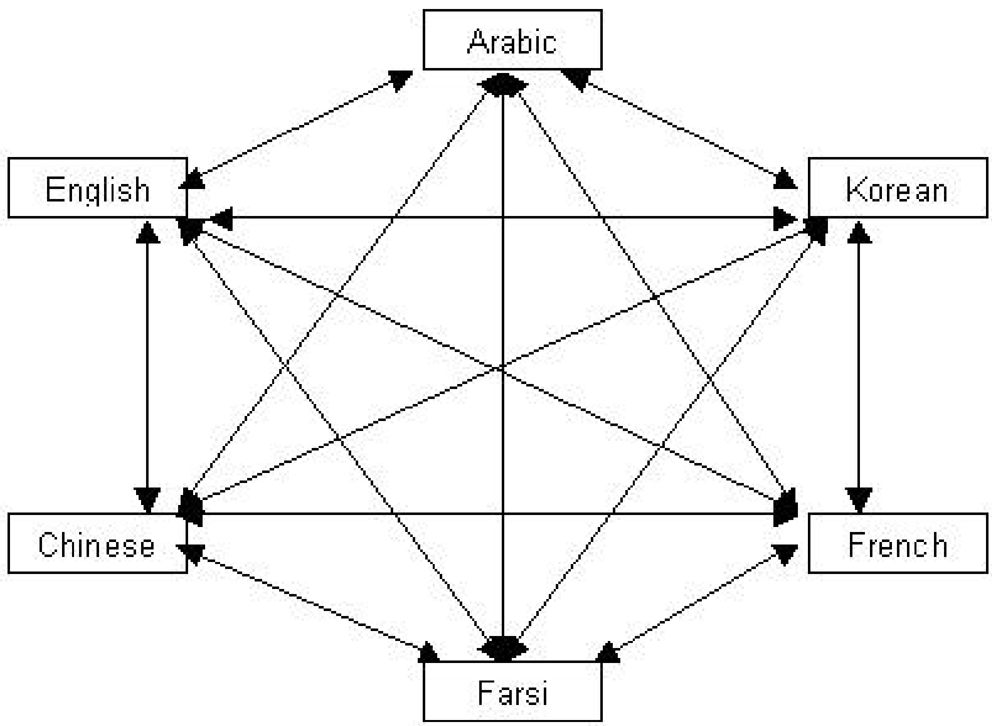
As shown in the above example, the Interlingua approach has two main stages of processing: the analysis of source language into Interlingua and the generation of target language from Interlingua.  Once the source language (SL) is analyzed into Interlingua representation, it can be mapped and generated into any target language (TL).  Therefore, it eliminates redundancy and simplifies the addition of other languages as well as results in high modularity.[35] Figure 1 depicts the advantage of the Interlingua approach.  Each language does not worry about the target language; however, the Interlingua approach requires a very rich and vast Interlingua representation to cover all phrases and words from all languages.  As such, it can be difficult to ensure that conversion always takes place consistently between each pair of languages.[36] For this difficulty, not many MT systems currently are able to incorporate an Interlingua approach into their system.




(note: each arrow depicts two directions: analysis and synthesis)

FIGURE1.  MULTILINGUALITY – INTERLINGUA

Lastly, the Transfer approach is a cross between the two previous methods.  The source language is converted into source language representation first and later into target language representation.  Here, each SL and TL pair has its own specific SL-TL representation.  This SL-TL representation is then synthesized into the target language, or transferred.  Although this is less efficient than the Interlingua approach, it has the advantage that the specific intermediary language is more specific to the languages being converted and it can therefore give better results. However, its disadvantages are that any modifications affect several transfer modules because of the specifically prescribed SL and TL representation relationship. Another disadvantage for transfer approach is its inefficiency for multilingual application. Each pair of languages requires its own specific SL and TL representation pair defined as shown in Figure 2. Every line between a pair of languages indicates two translations, source representation to target representation and target language representation to target language synthesis.




(note: each arrow depicts two directions: analysis and synthesis)

FIGURE 2. MULTILINGUALITY – TRANSFER APPROACH

Multilinguality translation may serve as a major distinction between the interlingua and the transfer approach. For example, Figure 1 and 2 shows six-language multilinguality case for both the interlingua and transfer approach. In the case of the interlingua, six source languages are translated (analyzed) into the interlingua (six total translations from source to the interlingua) and translate (synthesis) six target languages from the interlingua (six translations from the interlingua to six target languages), requiring 12 translation actions, or 2 times *N* where *N* is number of languages (*N* is 6 in this example). To provide the same translations, the transfer approach requires translating each direction for every pair of language and there are 15 combinations of pairs, requiring 30 translations for six languages to be communicative, or *N* times (*N*-1).

**Types of Machine Translation (MT)**

A typical MT system has two main components: a dictionary and a parser. The parser is used to analyze the source language and generate a parse of the contents. MT uses it to analyze a sentence and assign a description of syntactic structure with respect to grammar and lexicon. Words are assigned to certain categories and the structure is worked out using a parse tree. Semantic interpretation may take place later. Grammar and lexicon are used to provide the rules for assigning structure during parsing. The grammar contains grammatical categories that determine which combinations of certain types of words may belong to which larger category. The larger the grammar, the more capable the parser, but the slower it is. The lexicon is a database of words that provides information about which category the word may belong to, singular and plural forms and so on. Here, a simple sentence may be from a noun and a verb, such as, "The hunter catches a deer." The noun is "hunter" and the verb is "catches." So a parser uses rules like these to build a tree of the structure of the sentence. This may be done starting with the sentence and working to smaller categories, or it may be starting with individual words and working up to larger categories.[37]
There are mainly three types of Machine Translation (MT).

TRANSFER-BASED MT

Transfer-based MT performs analysis using a morphological analyzer, parser and grammar. Depending on the approach, the grammar must build either or both syntactic and or semantic representation to yield three kinds of transfer-based MT: Syntactic, Semantic, and Lexicalist. Syntactic MT rearranges phrases and translates lexical items; it is also a relatively easy program to write. Semantic MT offers the greatest chance of meaning preservation during the translation and has simpler transfer rules. The Lexicalist MT offers transparent transfer rules and is less theory dependant. Translation equivalence between sets of lexicons is easier to verify.[38] Transfer-based MT process can be seen in Figure 3.

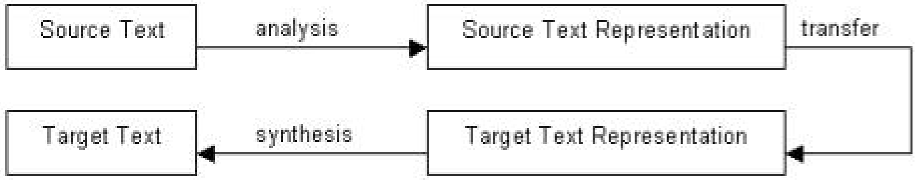



FIGURE 3. TRANSFER-BASED MT PROCESS

### INTERLINGUA-BASED MT

Interlingua-based MT performs analysis using a parser and possibly a separate semantic interpreter. Unlike the Transfer-based MT, the interlingua based MT does not have the intermediate process such as source text and target text representation as shown in Figure 4.

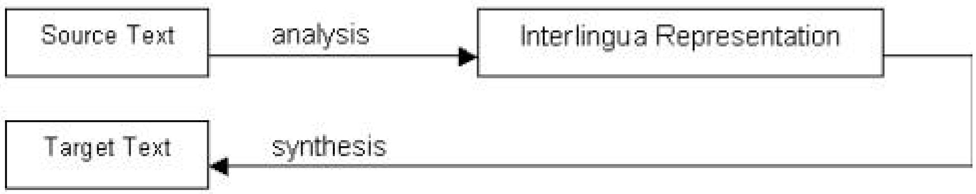



FIGURE 4. INTERLINGUA-BASED MT PROCESS

There are two types of interlingua-based MT: 1) Linguistics-based and 2) Knowledge-based MT. Linguistics approaches are mainly based on syntactic patterns and constraints with the meaning representation providing sufficient basis for an Interlingua representation. On the other hand, linguistic meaning is dependent on non-linguistic knowledge in the Knowledge-based Interlingua. It uses real world knowledge to augment meaning representations. World and domain knowledge is useful for handling ambiguity, but its keen domain dependency requires complex analysis and generation.[39]

### EXAMPLE-BASED MT (EBMT) AND TRANSLATION MEMORY (TM)

EBMT and TM are the latest developments of MT leveraging computer technology. Example-based MT, also known as Corpus-based, is essentially translation by analogy from example. An EBMT system is given a set of sentences in the source language and their corresponding translations in the target language, and then uses those examples to translate other similar source-language sentences into the target language. The basic premise is that if a previously translated sentence occurs again, the same translation is likely to be correct again. EBMTs are portable to new domains and language pairs. They are more extensible than rule-based systems. Some EBMT systems extract translation patterns or templates from bilingual text. The biggest problem the EBMT system faces is that it needs large amounts of pre-translated text examples to make a reasonable general-purpose translator. To make the use of examples more effective, example databases can be generalized so that more than one input string can match any given part of the example.[40]

The EBMT process is divided into the three tasks of matching source language fragments of an input against a database of translation examples. It identifies the corresponding target language fragments and then combines them appropriately to produce a target language string. These steps can be illustrated by means of a Vauquois triangle with the tasks of EBMT superimposed in the pyramid in Figure 5.[41]

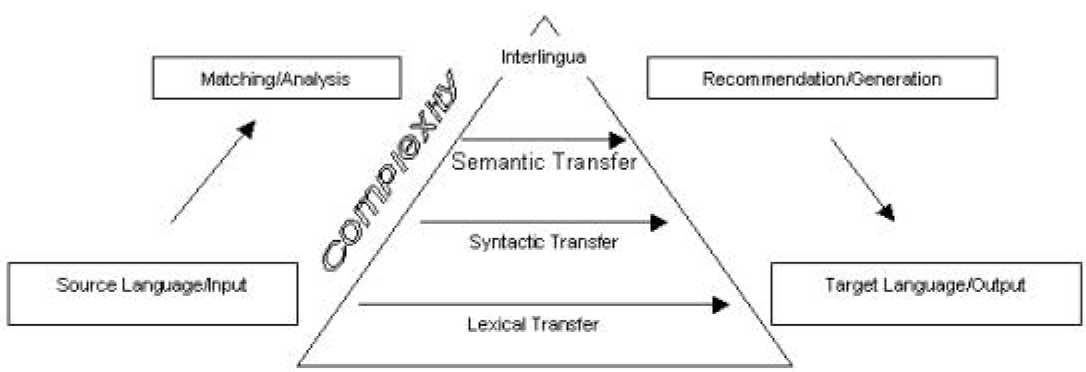



FIGURE 5. VAUQUOIS TRIANGLE WITH EBMT PROCESS[42]

Translation Memory is a restricted form of example-based translation taking advantage of computing power. Many recent commercial MT systems have TM as part of the system. In a translation memory, as the user translates text, the translations are added to a database, and when the same sentence occurs again, the previous translation is inserted into the translated document. This saves the user the effort of re-translating that sentence, and is particularly effective when translating a new revision of a previously translated document.[43] For this reason, Example-based and TM-based MT systems can greatly improve their translation qualities by developing large example databases, or corpus. Developers of both systems constantly look for parallel texts in many different domains to improve and expand the current capabilities. Of all the MT approaches and systems, Example-based and Translation Memory seem to have the most potential for the short- term as those databases can rapidly expand through the Internet.

## TYPES OF USE

Jarmie Carbonell of Carnegie Mellon University categorized, as shown in Figure 6, the functional types of translations as mainly dissemination and assimilation, and suggested that the dissemination side would require much higher translation quality as compared to that of the assimilation. The basic reasoning is that the text on the dissemination side contains specific information to be shared with the reader, whereas in assimilation, the kind of information to be extracted is largely dependent on the specific interest of the reader. This is an important distinction for developing a specific corresponding MT system for a specific application as it would significantly increase the overall translation quality.

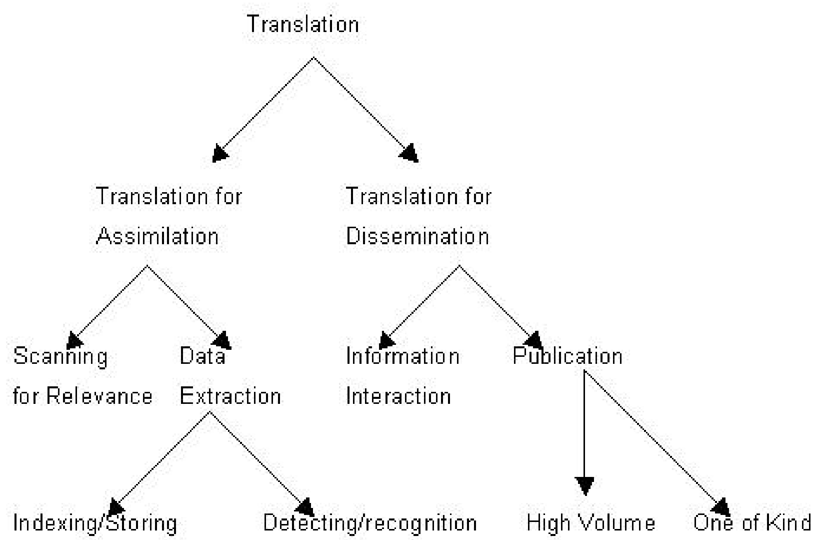



FIGURE 6. TYPE OF MACHINE TRANSLATION USAGE[44]

## CHALLENGES

Translating between languages is complex even for humans. The best translations are not simple word-for-word substitutions. In a famous example, “Out of sight, out of mind” translates to “invisible idiot.”[45] An ultimate translation captures the intended core meanings and transposes them into knowledge in the other language. Implementing this process is not an easy matter, requiring tremendous effort to give the computer the knowledge or set of rules it needs to translate correctly. The assumption in MT systems, whether fully or partially automatic, is that there are sufficiently large areas of natural language and translation processes that can be formalized for treatment by computer programs. Therefore, the basic premise is that the differences between languages can to some extent be regularized. What this means at the practical level is that problems of selection can be resolved by clearly definable procedures. The major task for MT researchers and developers is to determine what information is most effective in particular situations, what kind of information is appropriate in particular circumstances, and whether some data should be given greater weight than others.

Perhaps the most challenging issue for MT is how to resolve ambiguity, homonymy, and alternative structure. In many instances, a same word can have different meanings depending on context. The issue of ambiguity occurs in every step of the MT process, in the analysis of the source text, the bilingual transfer of lexical items and structure, and the generation of the target text. If the disambiguity process fails during any of the three stages, output of the MT would not be good – a “garbage in, garbage out” type of process.

One effective way the MT community has dealt with these challenges is to use controlled language: limit the amount of choices in the actual texts input to the MT system or to limit the system itself, text types, or subject areas. It also requires texts to conform to certain restrictions of vocabulary and syntax with a specific set of rules. This is the process of matching the MT system to a specific task or domain.

New MT approaches, specifically Example-based and Translation Memory seem very encouraging. It uses parallel databases which contain the same translated sentences in the source and target languages. In addition, it uses bilingual dictionaries and special algorithms, including some statistical techniques, to match-up corresponding words and phrases of the sentences in the two languages. The computer remembers these matches and, when presented with a new sentence, retrieves the matches and pieces them together to produce a translation for the new sentence. This empirical approach to machine translation is becoming more popular because it requires less human effort and can produce a working system in less time. In addition, the technology and data resources needed to develop it are constantly improving. The resulting systems can have a level of performance that approaches that of "knowledge-based" systems for significantly less cost.

COMMERCIAL MT SYSTEM ASSESSMENT

U.S. Joint Forces Command's Joint C4ISR Battle Center (JBC) conducted a comprehensive assessment on a Commercial-Off-the-Shelf (COTS) Machine Translation product entitled Text Simultaneous Machine Translation Assessment Report in October 2002. Previously, all combatant commanders raised their concerns about the deficiencies of foreign language capabilities in their commands. This assessment was a way to investigate if any COTS products are mature enough to be used as a tool to augment the shortfalls. One common recommendation the JBC made in their report on the use of MT systems for combatant command was a lack of a concept of operations (CONOPS). Introduction of any new capabilities to the operating forces will require a CONOPS and Tactics, Technique and Procedure (TTPs) to support employment of the capabilities as noted in the previous section. The JBC report concluded that Machine Translation is a viable tool to support the warfighter today.[46] Details of the assessment will be omitted here because it rank orders commercial products.

APPLICATIONS

Machine Translation systems, like many commercial technologies, are applicable to the national and military challenges noted earlier in this paper. Their use can reduce system costs and improves utility. There are, however, obligations concomitant with their use, just like all other uses of COTS. Often, commercial technologies accompany commercial practices. Applications of COTS or any tools for new tasks require some level of preparation such as developing training or writing new operating procedures in order to maximize the benefits and make efficient use of COTS. A single concept or device that will immediately produce the ascendancy of the user's forces over those of the user's adversaries does not work well with COTs. A revolutionary process is an evolutionary process in many ways. When an item in an evolutionary process achieves critical speed and mass, it can go revolutionary, and untie itself from the evolutionary orbit. A revolutionary process is like adding a drop into a glass filled with water to the top. Science, technology, and military inventions are all in need of such a progressive approach.

When Machine Translation (MT) was applied in 1970s and 1980s, peoples' expectations of MT were somewhat tame. To begin with, most of the public was not aware of any machine that could translate human language. In spite of its capability, MT was perceived by many people as a research and developmental gadget held hostage to the laboratory environment.

During the 1990s expectations have been much different. Technology growth has been phenomenal. Many families now own personal computers and are connected to the world by information technology. However, technological growth also has brought along its own sets of challenges and its own dilemmas. People begin to believe in and develop new views on technologies and its capabilities. The majority of the time, the belief is reasonably derived from reasonable assumptions. However, because information is moving through the Internet at an unprecedented speed, sometimes it is difficult to make sure what is presented is accurate. In some sense, people begin to believe that all of the technologies featured in Popular Science magazine or posted on websites work flawlessly without a glitch.

Reality is quite different. Even the most technologically advanced state-of-the art spacecraft in the world has to offer still needs fuel to operate. Machine Translation went through such hyped publicity in 1980s when a resurgence of interest made the headlines in Japan and the U.S. Many new consortia were formed as private companies and research organizations launched new ventures, trying to develop MT software and systems. Many began to believe in the technology. However, what they did not hear or the MT community failed to inform them was that MT is a tool, nothing more. It cannot possibly translate any “X” language to “Y” language in a perfect manner. It is not designed to handle that, and as a matter of fact the fundamental theories have not been fully developed. In the end it made many believers non-believers.

This is not to say that any technology which has not matured to 100 percent complete should be discarded. On the contrary, it should be used if there is an area where it can provide a utility. In order to move from an “evolutionary” to “revolutionary” tool, one has to think of and incorporate an engineering approach rather than a research approach. Utility of MT must be carefully assessed against requirements, asking what it can do and cannot do against a specific task. In many instances, an 80 percent solution is far better than no solution at all. It must be remembered though that a 20 percent shortfall in technology must be scrutinized before using it for future risk mitigation.

From this perspective, Machine Translation has a lot to offer, particularly for national and military operations. The current Internet is loaded with data as a result of incorporating HTML as its main language allowing it to easily place data on the web. One drawback is the time it takes finding the information requested from so much data. Machine Translation systems can provide great support in this area such as key word searching. Rather than human operators reading through one by one, MT can rapidly scan through the material and identify sections with key words or paragraphs.

The real benefit of machine translation would be in coalition military operations. During Operation Enduring Freedom (Afghanistan 2002), U.S. forces had to deal with indigenousness adversaries with 16 different nations' military forces, many of whom English was not their native language. Sharing and extracting information from locals was challenging and time-sensitive. U.S. military police detained and interrogated more than 3,000 detainees who did not speak or understand English. Although at the time of that operation the U.S. military had a number of linguistic specialists, it was still a very difficult task.

In many instances, there are standard sets of questions to be asked during an interrogation session. This process can be greatly assisted by machine translation. After all, the computer has an infinite amount of patience when it comes to repetitive tasks. At higher headquarter levels in which multinational coalition forces are working side by side, MT can be a tremendous tool to communicate the gist of meanings. For example, NATO or U.S.-ROK Combined Forces Korea would be an ideal candidate for machine translation application.

**MT for U.S.-Republic of Korea (ROK) Combined Force Command (CFC)**

U.S.-ROK CFC offers an ideal case for MT in military operations. The Republic of Korea's military command structure is very complex. Established in 1978, it is the combined warfighting headquarters for both the U.S and ROK. Throughout the command structure, bi-national manning is readily apparent: if the chief of staff section is filled by U.S. military personnel, the deputy and his staff will be Korean and vice versa. This integrated structure exists within the component commands as well as the headquarters. All CFC components are tactically integrated through continuous combined and joint planning, training, and exercises. CFC has operational control over more than 600,000 active-duty military personnel of all services, of both countries. In wartime, additional forces could include some 3.5 million ROK Reservists and U.S. forces based outside the ROK. U.S. augmentation forces are integrated into the appropriate CFC/USFK commands. Unity of command, therefore, is very crucial. For that reason, one U.S. general officer serves concurrently as the Combatant Commander of the multilateral United Nations Command (UNC), the bilateral U.S.-ROK CFC, and the U.S. Force Korea (USFK) command. The CFC and the UNC are legally separate military organizations. This UNC-CFC arrangement allows additional countries to send forces to the Korean Peninsula providing support to the UNC under operational control of Combatant Commander UNC while coordinating their operations with the Combatant Commander CFC.[47]

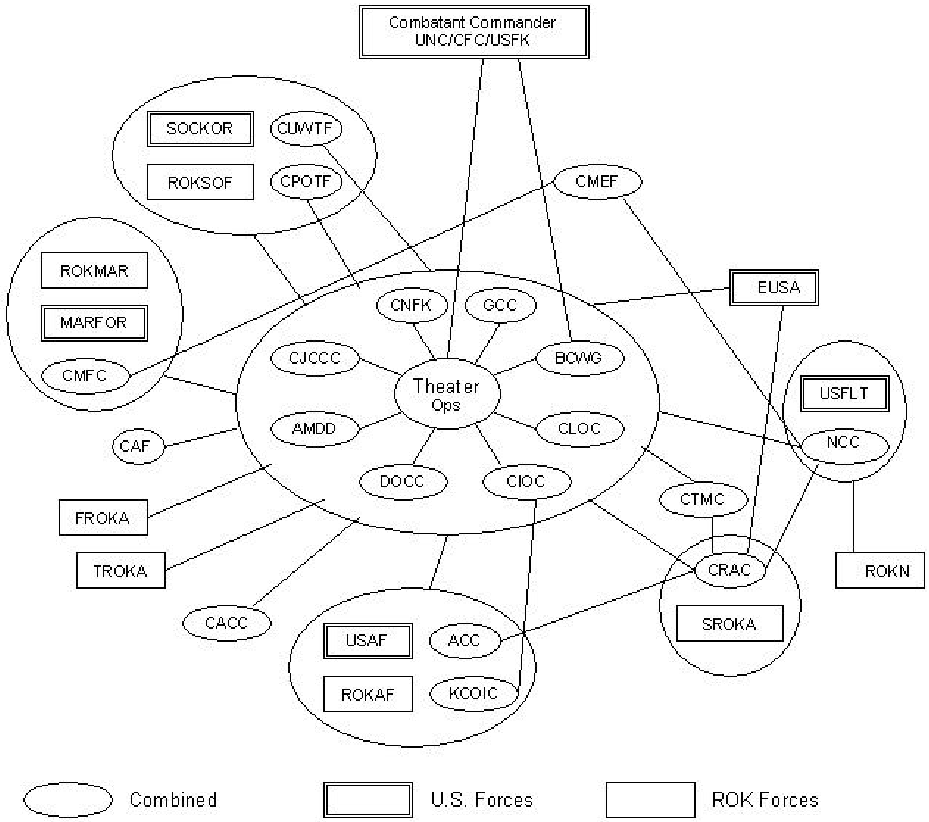



FIGURE 7. NOTIONAL CFC/USFK OPERATIONAL NODE CONNECTIVITY [48]

As shown, Figure 7 depicts the complexities for operating in a Combined and Joint environment. The figure also shows the lines of command and control (C2) between the various operational combined and national nodes. Each line represents the multiple exchanges that occur within the Theater Operations, all of which support C2 for the CFC commander and his forces. Information exchanges occur using various systems and communications media from Local Area Network (LAN) to Wide Area Network (WAN). This complexity poses interoperability challenges. From the U.S. point of view there are several operational factors that should be considered when addressing interoperability challenges.[49]

Interoperability solutions for UNC/CFC are often driven by a mixture of technology and policy. CFC is a Combined organization and is not staffed solely according to U.S. doctrine, which leads to different C4I infrastructures and business processes. Language translation and multi-level and multi-cultural security are major obstacles that information interoperability must overcome. Joint Publication 1-02 defines interoperability as "the ability of systems to provide services to, and accept services from, other systems and to use the services exchanged to enable them to operate effectively together."[50] However, from a war fighting point of view, interoperability involves more than ensuring systems can exchange information and operate effectively. Conducting a battle involves using information that may travel across multiple communications means and automated applications.

For the warfighters, the definition of interoperability is expanded to mean providing timely, accurate, and complete information at the right place and time to people who need it. In this light, interoperability challenges should be viewed as pieces of a puzzle. As each challenge is identified and solved, other challenges become evident as the result of information availability.[51] Figure 8 depicts a notional CFC objective C4I architecture for the FY05 to FY10 timeframe. It would support most of the currently identified interoperability challenges.

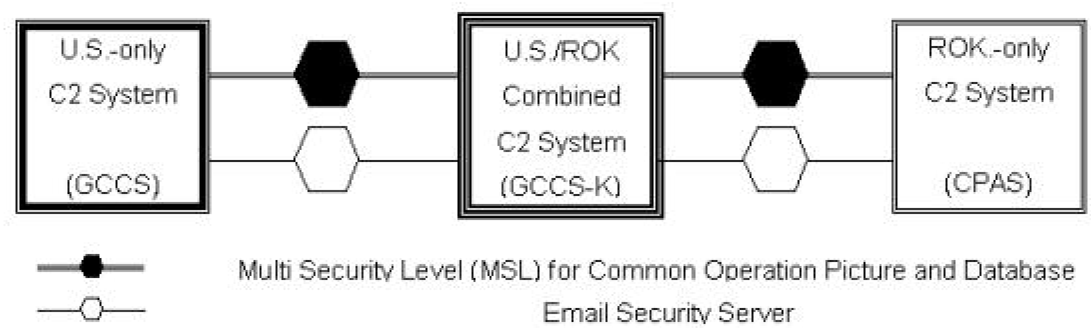



FIGURE 8. NOTIONAL CFC/USFK C4I ARCHITECTURE

This clearly begs the next question. Since digital bits and bytes connectivity do not render itself information or knowledge, what is the meaning of electronic connectivity such as e-mail between U.S.-ROK units and action officers when neither side has language understanding? In addition, what about the time of crisis when quick and swift information exchange and understanding is required for combined operations?

The new architecture needs to examine not only how the electrons flow from one to others, but it also needs to consider how the electrons can make sense to the recipients.

Currently, human translators are in great demand in UNC/CFC. They cannot be at every single terminal to decipher the plethora of incoming e-mails. A preferred, logical step is to install Machine Translation servers between or on the U.S. and ROK C2 systems so that those who receive e-mail can translate on-line.  E-mails and other similar documents such as graphical presentations are more often than not in non-standard English or Korean so that it would impact the quality of translation output.  However, since e-mail and graphical presentations are frequently used to convey concise messages, perhaps the gist of the message would be very useful and not too difficult to be captured by translation machine.  Figure 9 shows a modified notional C4I architecture in which machine translation capabilities are incorporated to provide such utilities.

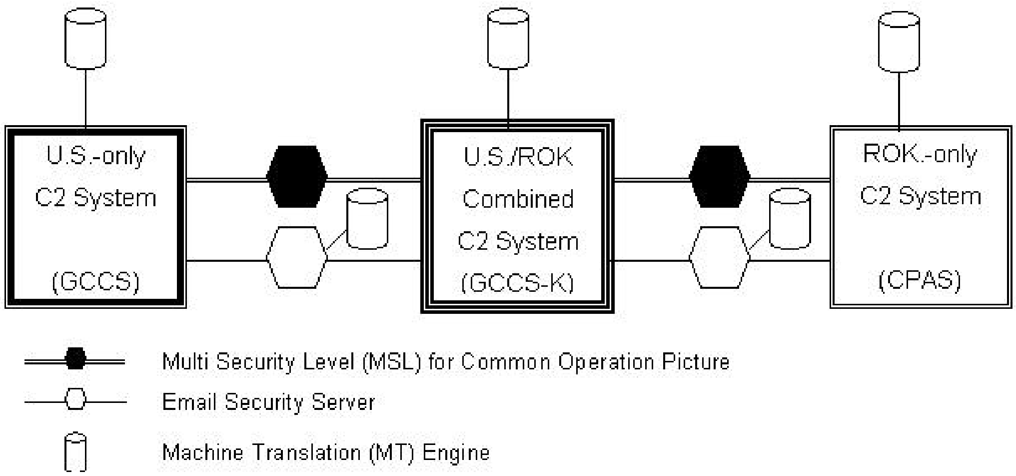



FIGURE 9.  NOTIONAL CFC/USFK C4I ARCHITECTURE WITH MT CAPABILITIES

**CONCLUSION**

The Internet is all about information, and more specifically how to effectively access, manage, understand, and turn that into one's own knowledge and power.  Although the Internet takes people to the information gate, one has to navigate through many different portals to find what one is looking for.  Information can be in any language – about 60 percent of data posted on the Internet is non-English and the ratio is growing.  There are simply too many different languages, and each of them is important to somebody.  From the national security and interest point of view, competition in the information race intensifies as other nations now have the same access to the information gate.  Who gets the information first matters, and who exercises the power of information first matters to national security.  Foreign language capabilities play a key role in this race.  However, the current level of capability is inadequate in meeting the national requirement.  It will take serious investments, time, and planning to establish the desired number of linguists with foreign language skills.

Machine Translation technology can serve as an interim solution for the shortfall of human translators by providing sound augmentation. While the technology is not as robust as some may perceive it, through an innovative engineering approach it can surely help human translators or users to perform their task better and faster. It is a tool to aid human activities, not to replace humans. Therefore, it is very important to recognize its capabilities of what it can and cannot do. Applying the military's target-weapon pairing approach would maximize MT utilities for specific applications such as email and Internet websites. As information technology flourishes, the demand for immediate translations will continue to grow rapidly and eventually provide a seamless integration of information. The technology is growing rapidly and in a short time MT will be an integral part of a true human-centric system which is a key to Information Supremacy and Knowledge-based operations.

WORD COUNT = 6953

## ENDNOTES

[1] International Bible Society, The Holy Bible: New International Version (Michigan: Zondervan Co., 1982), 9-10.

[2] The American Heritage Dictionary: 2d College ed. (Boston: Houghton Mifflin Co., 1976), 705.

[3] Definition by Koh: *knowledging* is a proactive and conscious action or process being taken to acquired knowledge and is different from the conventional definition of information processing. From this sense "information overload" is not possible. The human brain far exceeds processing power of any computer. If the humans are overloaded, it is because information provided is not pertinent and relevant. If this is the case, then what is being provided is not information, but raw data. Data is, in my view, unprocessed and unprepared. It must be processed within a particular context. A good example of this is data on the WWW or Internet. It is up to the finders or users to give or assign the relevancy to data: someone's garbage is another's treasure.

[4] Mike Urban, "Esperanto," 23 June 1999; available from <http://www.esperanto.net/veb/faq.html>; Internet; accessed 23 January 2003.

[5] Thomas L. Friedman, The Lexus and the Olive Tree, (New York: Anchor Book Co., 1999), 7-8.

[6] Roland Robertson, Globalization: Social theory and Global Culture, (London: Sage,1992), 8.

[7] Martin Albrow, The Global Age, (London: Sage, 1996), 88.

[8] Ronald Cole et al., Survey of the State of the Art in Human Language Technology, (Cambridge: Cambridge University Press, 1996), 1-5.

[9] John Hutchins, "Why computer do not translate better," Journal of Translating & the Computer 13 (November 1992): 5-6.

[10] Congress, Senate, Committee on Government Affairs, Subcommittee on International Security, Proliferation, and Federal Services, The State of Foreign Language Capabilities in National Security and the Federal Government: Hearing before the Subcommittee on International Security, Proliferation, and Federal Services, 106th Cong., 2d sess., 14 and 19 September 2000, 1.

[11] Ibid.

[12] Ibid., 8.

[13] General Accounting Office, FOREIGN LANGUAGES: Workforce Planning Could Help Address Staffing and Proficiency Shortfalls (Washington, D.C.: U.S. General Accounting Office, March 12, 2002), 1.

[14] Ibid., 8.

[15] Congress, Senate, Committee on Government Affairs, Subcommittee on International Security, Proliferation, and Federal Services, The State of Foreign Language Capabilities in National Security and the Federal Government: Hearing before the Subcommittee on International Security, Proliferation, and Federal Services, 106th Cong., 2d sess., 14 and 19 September 2000, 3-5.

[16] General Accounting Office, FOREIGN LANGUAGES: Workforce Planning Could Help Address Staffing and Proficiency Shortfalls (Washington, D.C.: U.S. General Accounting Office, March 12, 2002), 2.

[17] Congress, Senate, Committee on Government Affairs, Subcommittee on International Security, Proliferation, and Federal Services, The State of Foreign Language Capabilities in National Security and the Federal Government: Hearing before the Subcommittee on International Security, Proliferation, and Federal Services, 106th Cong., 2d sess., 14 and 19 September 2000, 2.

[18] Ibid., 9.

[19] Carl Stiner et al., Report of the Senior Working Group on Military Operations Other than War (Arlington: US Advanced Research Project Agency, May 1994), 11.

[20] Congress, Senate, Committee on Government Affairs, Subcommittee on International Security, Proliferation, and Federal Services, The State of Foreign Language Capabilities in National Security and the Federal Government: Hearing before the Subcommittee on International Security, Proliferation, and Federal Services, 106th Cong., 2d sess., 14 and 19 September 2000, 61.

[21] Ibid., 2.

[22] General Accounting Office, FOREIGN LANGUAGES: Workforce Planning Could Help Address Staffing and Proficiency Shortfalls (Washington, D.C.: U.S. General Accounting Office, March 12, 2002), 7.

[23] Jonathan Slocum, "A survey on Machine Translation: Its History, Current Status and Future Prospects," Computational Linguistics, 11, no. 1 (1985): 1.

[24] The European Association for Machine Translation, "What is Machine Translation?" available from <http://www.eamt.org/mt.html>; Internet; accessed 10 Jan 2003.

[25] John Hutchins, "Machine Translation and Human Translation: In competition or in complementation?" International Journal of Translation 13, no.1-2 (2001): 13-14.

[26] Walther von Hahn, "Innovative Concepts for Machine Aided Translation," Proceedings of VAKKI (1992): 15.

[27] Martin Kay, “Machine Translation,” available from <http://www.lsadc.org/Kay.html>; Internet; accessed 15 Jan 2003.

[28] Ibid.

[29] Slocum, “A survey on Machine Translation: Its History, Current Status and Future Prospects,” Computational Linguistics, 11, no 1, (1985): 1.

[30] Ibid.

[31] John Hutchins, “Machine Translation and Human Translation: In competition or in complementation?” International Journal of Translation 13, no.1-2 (2001): 14.

[32] Martin Kay, “Machine Translation: The Disappointing Past and Present,” available from <http://cslu.cse.ogi.edu/HLTsurvey/ch8node4.html>; Internet; accessed 20 December 2002.

[33] Joanne Birt, “Natural Language Processing,” available from <http://www.comp.leeds.ac.uk/ugadmit/cogsci/spchlan/nlp.htm>; Internet; accessed 20 December 2002.

[34] Ibid.

[35] Arturo Trujillo, “Strategy for Machine Translation,” available from <http://www.ccl.umist.ac.uk/staff/iat/transeng/barca.ppt>; Internet; accessed 21 December 2002.

[36] Joanne Birt, “Natural Language Processing,” available from <http://www.comp.leeds.ac.uk/ugadmit/cogsci/spchlan/nlp.htm>; Internet; accessed 20 December 2002.

[37] Ibid.

[38] Arturo Trujillo, “Strategy for Machine Translation,” available from <http://www.ccl.umist.ac.uk/staff/iat/transeng/barca.ppt>; Internet; accessed 21 December 2002.

[39] Ibid.

[40] Jaime G. Carbonell, “Example-Based Machine Translation,” available from <http://www-2.cs.cmu.edu/~ralf/ebmt/intro.html; Internet; accessed 22 December 2002.

[41] Kevin McTait, Translation Pattern Extraction and Recombination for Example-Based Machine Translation: PhD Thesis. (Manchester, UK: University of Manchester, September 2001), 8-13.

[42] B.A. Vauquois, “Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation,” IFIP Congress-68 (1968): 254-260.

[43] Jaime G. Carbonell, “Example-Based Machine Translation,” available from <http://www-2.cs.cmu.edu/~ralf/ebmt/intro.html; Internet; accessed 22 December 2002.

[44] Walther von Hahn, “Translation Technology,” available from <http://www.racai.ro/awd/awd16/hahn.html>; Internet; accessed 27 December 2002.

[45] John Hutchins, “Machine Translation and Human Translation: In competition or in complementation?” International Journal of Translation 13, no.1-2 (2001): 5-20.

[46] U.S. Joint Forces Command Joint, Text Simultaneous Machine Translation Assessment Report, BC JFCOM Report, (Suffolk: Joint Forces Command, October 2002), 4-2.

[47] MITRE Corporation, CFC/USFK C4ISR Interoperability Challenges, (McLean:MITRE, August 2000), 3.

[48] Ibid., 16.

[49] Ibid., 15.

[50] Ibid., 38.

[51] Ibid.

## BIBLIOGRAPHY


Albrow, Martin. The Global Age. London: Sage, 1996.

Birt, Joanne. “Natural Language Processing.” Available from <http://www.comp.leeds.ac.uk/ugadmit/cogsci/spchlan/nlp.htm>. Internet. Accessed 20 December 2002.

Brown, Ralf D. “Automated Generalization of Translation Examples.” Proceedings of the 18th International Conference on Computational Linguistics (July 2000): 125-131.

Carbonell, Jaime G. “Example-Based Machine Translation.” Available from <http://www-2.cs.cmu.edu/~ralf/ebmt/intro.html>. Internet. Accessed 22 December 2002.

Carbonell, Jaime G. “Is There Natural Language after Data Bases?” Proceedings of 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics (July 1984): 186-187.

Cole, Ronald et al., eds, Survey of the State of the Art in Human Language Technology. Cambridge: Cambridge University Press, 1996.

Friedman, Thomas L. The Lexus and the Olive Tree. New York: Anchor Book Co., 1999.

Furu, Sadaoki. “Automatic Speech Recognition and Its Application to Information extraction.” Proceedings of 37th Annual Meetings of the Association for Computational Linguistics (July 2000): 11-20.

Hahn, Walther von. “Innovative Concepts for Machine Aided Translation.” Proceedings of VAKKI (1992): 13-25.

Hahn, Walther von. “Translation Technology.” Available from <http://www.racai.ro/awd/awd16/hahn.html>. Internet. Accessed 27 December 2002.

Halteren, Hans van, Walter Daelemans, and Jakub Zavrel. “Improving Accuracy in Word Class Tagging through the Combination of Machine Learning Systems.” Computational Linguistics 27, no. 2, (June 2001): 199-230.

Hutchins, John. “Machine Translation and Human Translation: In competition or in complementation?” International Journal of Translation 13, no.1-2 (2001): 5-20.

Hutchins, John. “Why computer do not translate better.” Journal of Translating & the Computer, no.13 (November 1992): 3-16.

International Bible Society. The Holy Bible: New International Version. Michigan: Zondervan Co., 1982.

Karov;Yael, and Shimon Edelman. “Similarity-based Word Sense Disambiguation.” Computational Linguistics Special-Issue-on-Word Sense Disambiguation, Computational Linguistics 24, no. 1 (March 1998): 41-60.

Kay, Martin. “Machine Translation.” Available from <http://www.lsadc.org/Kay.html>. Internet. Accessed 15 Jan 2003.

Kay, Martin. “Machine Translation: The Disappointing Past and Present.” Available from <http://cslu.cse.ogi.edu/HLTsurvey/ch8node4.html>. Internet. Accessed 20 December 2002.

King, Margaret. “When is the next ALAPC Report Due?” Proceeding of 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics, (July 1984): 352-353.

Li, Cong, and Li Hang. ”Word Translation Disambiguation Using Bilingual Bootstrapping.” Proceedings of 40th Annual Meetings of the Association for Computational Linguistics (July 2002): 343-351.

McTait, Kevin. Translation Pattern Extraction and Recombination for Example-Based Machine Translation. PhD Thesis. Manchester, UK: University of Manchester, September 2001.

MITRE Corporation. CFC/USFK C4ISR Interoperability Challenges: Report to Commander-In-Chief Interoperability Office. McLean, VA, August 2000.

Robertson, Roland. Globalization: Social theory and Global Culture. London: Sage, 1992.

Slocum, Jonathan. “A survey on Machine Translation: Its History, Current Status and Future Prospects.” Computational Linguistics 11, no. 1 (1985): 1-12.

Stiner, Carl et al., Report of the Senior Working Group on Military Operations Other than War. Arlington: U.S. Defense Advanced Research Project Agency, May 1994.

The American Heritage Dictionary. 2d College ed. Boston: Houghton Mifflin Co.,1976.

The European Association for Machine Translation, “What is Machine Translation?” Available from <http://www.eamt.org/mt.html>. Internet. Accessed 10 Jan 2003.

Towel, Geoffrey and Ellen M.Voorheest. “Disambiguating Highly Ambiguous Words.” Computational Linguistics 24, no. 1 (March 1998): 125-146.

Trujillo, Arturo. “Strategy for Machine Translation.” Available from <http://www.ccl.umist.ac.uk/staff/iat/transeng/barca.ppt>. Internet. Accessed 21 December 2002.

U.S. Congress. Senate. Committee on Government Affairs. Subcommittee on International Security, Proliferation, and Federal Services. The State of Foreign Language Capabilities in National Security and the Federal Government: Hearing before the Subcommittee on International Security, Proliferation, and Federal Services. 106th Cong., 2d sess., 14 and 19 September 2000.

U.S. General Accounting Office. FOREGING LANGUAGES: Workforce Planning Could Help Address Staffing and Proficiency Shortfalls. Washington, D.C.: U.S. General Accounting Office, March 12 2002.

U.S. Joint Forces Command. Joint, Text Simultaneous Machine Translation Assessment Report. Suffolk, VA: Joint Forces Command, October 2002.

Urban, Mike. “Esperanto.” Available from <http://www.esperanto.net/veb/faq.html>. Internet. Accessed 23 January 2003.

Vauquois, B.A. “Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation.” IFIP Congrfess-68 (1968): 254-260.