A New Dawn for Machine Translation?
Reinhard Schäler

Machine Translation: An Imperfect But Evolving Technology
Lorena Guerra

The Case for Proofing 100% Matches
Hélène Wimmerlin

Translation Speed vs. Content Management
Jeff Allen

A Translator Protests Poor Use of Machine Translation
Carlo Capussotti

The Electric Word and the Semantic Web
Jaap van der Meer
Perspectives on Machine Translation

Machine translation (MT) is a topic that arouses a variety of passions, not only technical but also political and personal. In preparing issue #62, we received so much opinion and information about MT that it developed into a complex point-counterpoint with technical discussion. Aware that the technology and its context are changing rapidly, we created this special supplement to make these comments and ideas available to readers of MultiLingual Computing & Technology in a timely manner.

Jeff Allen introduces this supplement with some thoughts and questions to keep in mind while reading the various articles. He also contributes an article examining translation speed in contrast with content creation and management and illustrates ways of developing statistics to measure qualitative and quantitative change over time.

Reinhard Schäler questions some of the current claims being made for new machine translation tools and the long-promised dawn of MT.

Lorena Guerra offers the results of her tests in regard to whether machine translation with post-editing can save translators time in their process.

Hélène Wimmerlin points out the need for a translator or proofreader to review even text that is supposed to be a 100% match in order to achieve an accurate translation.

Carlo Capussotti relates some experiences and warnings regarding inappropriate or poorly planned use of MT and the consequences for subsequent review and corrections.

Jaap van der Meer concludes with a look beyond current technology toward the Semantic Web and what that may mean to people who have worked with words in other ways.

And in Resources we include links to other information sources.

Reader comments are always welcome.

— Laurel Wagers, Managing Editor

Jeff Allen is project manager and knowledge manager of the Mycom International Software Product Division. He is also a member of the MultiLingual Computing & Technology editorial board. He can be reached at jeff.allen@free.fr

Carlo Capussotti is a translator and QA specialist with Novilinguists Multimdia in Italy. He can be reached at carloc@novilinguists.com

Lorena Guerra is a Spanish translator and posteditor. She can be reached at lorena.guerra2@mail.dcu.ie

Reinhard Schäler is director of the Localisation Research Centre at the University of Limerick. He can be reached at reinhard.schaler@ul.ie

Jaap van der Meer published Electric Word and Language Technology magazine from 1987 to 1990. He is a partner in Cross Language N.V., a consultancy and systems integration company specializing in taxonomy, globalization management software and MT implementations. He can be reached at jaap.vandermeer@crosslang.com

Hélène Wimmerlin is an account executive for Iverson Language Associates, Inc., and associate lecturer in the translation program at the University of Wisconsin-Milwaukee. She can be reached at helene@iversonlang.com

This guide is published as a supplement to MultiLingual Computing & Technology, the magazine about language technology, localization, Web globalization and international software development.
Thinking About Machine Translation

Jeff Allen

As I read over the language technology focused articles presented in this supplement, I noticed that several points of debate have been raised. This overview offers a few questions to highlight points that I think readers should think about.

1. Defining Quality

There has been discussion about good versus bad translation quality, yet I would like readers to think about whether translation quality is a simple binary formula, if there is a spectrum of levels, and/or if one can easily determine the correctness of a translation. Is “translation” a single-type and single-approach process, or are there some approaches that are more appropriate or less appropriate for different types of translation and needs?

The translators and translation agencies whom I know all claim that they produce high-quality translations. The word quality in this field has various meanings depending on what it applies to. The concept of the word quality is relative to customer requirements and expectations, the time constraints of the project, the target register of language and style, technical accuracy and so on. Geoffrey Kingscott published a good summary article, “Translation Quality Assessment: An Overview,” in the May 20, 2003, issue of the LISA newsletter. Translation quality and appropriateness can only be determined by the evaluation of a translated product (text or voice) based on the input criteria and factors, not simply by the surface presentation of products.

2. Developers’ Promises

One of the articles addresses the promises made by MT developers. It is true that some promises in the past by MT promoters have been utopian, such as MT with no need of post-editing (MultiLingual Computing & Technology #6 Volume 5 Issue 2), which resulted in a change in perspective a few years later on the same project. Yet, have these promises in the past been only marketing and sales oriented, or was a part of the responsibility borne by the user institutions that possibly inadequately identified, analyzed, and presented their real needs to the MT companies? What about cases where an MT user institution purchases a tool for one type of need such as inbound content gisting but then later decides to use it for a completely different need such as an outbound translation for publication? Who is really at fault in such a situation — the tool vendor, the customer or both?

3. Source of Failures

Have the failures of MT implementation been solely due to the software/systems, or were there external factors and third-party application problems? Was task planning and human resource allocation adequate and appropriate for such projects? Were there issues of implemented third-party tools that had not been part of the original strategy, or were changes made along the way that substituted some external tools for others and thus resulted in system incompatibility issues?

4. Integration

Deploying language technology systems is just one step. Integrating them into task workflows with human translators requires good training that helps transition the users from the previous workflow to the new one. Have there been adequate training courses on MT over the years? If they have existed, did the MT user organizations order such training for their own users, or did they decide to do without such training and invent the user training and mentoring process themselves? In such conditions, who is at fault for the failure of the project?

5. Customer Knowledge

Should we expect inbound-translation content-gisting users to know the source language and choose the appropriate MT engine whereas automatic language identification techniques have been developed, tested and publicly presented over the past several years? Content gisting assumes that the users don’t know how to read the source language, so should we even expect them to know what language it is? This is important, given that only 1,000 or so of the 6,500 languages in the world today have a written form. Consider automatic language identification work by Philip Resnik (www.umiacs.umd.edu/~resnik/strand) and by AppTek (www.apptek.com).

6. Tracking Statistics

We have seen one article on the comparison of translation speed via two different translation workflow channels. Yet, how many translation departments and translation agencies keep track of translation statistics which can be used to benchmark their human-tasks productivity against translation software implementation efforts? How are translation organizations keeping track of their translation productivity?

7. Cause or Effect?

Are translation software providers creating products that integrate into an existing workflow process, or are they creating products that completely modify the existing one? How are current experienced translators involved in defining the specifications of the software and evaluating it?

8. Semantic Web

Is the Semantic Web with proposals of XML-based metadata schemes something that can really be achieved in the next few years, or are the data sources so diverse and nonstandardized that such a large-scale effort will be very slow? Also, is the mention of XML really an added value today, or is it just a marketing idea and sales pitch?

I hope that these questions will give readers something to think about. Ω
A New Dawn for Machine Translation?

Reinhard Schäler

Imagine that you have worked for decades on a project that promised to deliver nothing less than one of humanity's dreams. Suddenly you begin to realize that, despite all of your efforts, you will not be able to deliver. Imagine — and this would surely be an even more difficult situation — that you have already announced that your efforts to realize this dream have been successful when, in reality, you know that you are going to fail miserably. What are your options?

According to a paper recently published in the proceedings of ASLIB, Translating and the Computer 25, by one of the most respected and eminent industry figures, Jaap van der Meer, “Machine translation (MT) has long been a controversial topic, the source of illusions, jokes and even serious disputes. Research and development in fully automatic translation have been carried out for fifty years. At regular intervals, researchers...have heralded the big breakthrough.”

Van der Meer is right. MT developers announced again and again that they had finally proven that their fully automated high-quality MT systems could make humanity's dream a reality. They claimed to have solved the dilemma created by the Tower of Babel with the help of a clever computer program.

Because executives with purchasing power (and little understanding of the issues involved) were easily impressed by “hard figures,” it became fashionable to randomly quote percentage values as a measure of success in MT — 95% accuracy, a minimum of 40% savings, 100% consistent use of terminology and so forth.

Blinded by the enormous potential savings, these executives bought high-powered MT systems and employed linguists to fine-tune and program them. This managed to create short-lived, cyclical surges of interest (and investment) in MT. However, they were always soon followed by a depression (and redundancies) because customers quickly realized that, once again, overenthusiastic sales executives had taken them for a ride.

At ASLIB, van der Meer announced that this time “the breakthrough is true” and that it is “market-driven rather than technical.” He conceded that “MT is not perfect” but that “it has become an economic necessity.” He presented figures to prove that, finally, MT can be employed in such a way that it makes economic sense.

Another respected professional, Ross Smith from PricewaterhouseCoopers, provided “An Overview of PwC/Systranet Online MT Facility” at the same conference, offering use statistics and feedback in addition to some practical examples around areas such as gisting.

The user feedback reported by Smith is positive overall. (Frankly, looking at the ever-so-slightly-skewed questionnaire, one cannot entirely discard the suspicion that the positive user response was at least partially preprogrammed.) The only negative example reported came, to the amusement of the presenter and his audience, from a user who had attempted to translate a document written in Dutch into English, using a German→English MT engine. No wonder he was not happy with the translation!

So, it seems that there is not just one, but many reasons to push the boat out and organize celebrations for the new dawn of MT, the final breakthrough.

Or are there?

It took smart MT developers years to comprehend that, even in a sales situation, the correct answer to the question “Will MT help us to save money on translation?” is “It depends.” They realized that although this is probably the most difficult answer of all, the alternative, that is, quoting misleading and simplified percentage figures, was not appropriate and had caused, on previous occasions, much damage to the MT community, undermining its reputation and credibility.

Not only are we now presented with yet another set of figures intended to prove the success of MT, but also with this new vision that it is okay if MT produces bad translations. Apparently, the correct use of language does not matter as long as it makes good business sense: MT is really our only hope to deal with the enormous volume of material that needs to be translated in our globalized economy.

Most US programmers understood, even in the early days, that people speak different languages and that there was a need for what became known as localization. It just took linguists some time to explain to them that even little marks on individual letters (accents and the like) can make a significant difference in the meaning of otherwise identical words. Finally, programmers agreed that 7-bit ASCII was not sufficient to represent meaning in languages other than English. The day they did was truly a day for celebration.

Are we now being told that language rules, syntax, grammar and terminology (never mind little marks on top of letters) do not really matter any more because there is a business imperative to use MT?

We are told that MT is most successfully being used not by language professionals, but by uninstituted office workers who just want to get the gist of a document that has been sent to them in a language they do not understand.

When it transpires that they are not even able to determine the language that the original document was written in and therefore they pick the wrong MT engine, it does not raise a red flag in our mind but, rather surprisingly, just causes the expert and his audience to pity the poor simple-minded user.

Listening to the experts, as a user and researcher, one cannot but sense that this final “breakthrough” has just pushed the bottom out of the barrel. It appears that the red light we see on the horizon is the sun finally setting on MT, rather than the announcement of a new dawn for the realization of one of humanity’s long cherished dreams.
**Machine Translation: An Imperfect But Evolving Technology**

Lorena Guerra

Fully automatic high-quality machine translation (FAHQMT) was an ideal for many years for developers and implementers of machine translation (MT) systems. The aim was to produce high-quality translations of natural languages without any human intervention, except, of course, clicking on the “Translate” button. Nevertheless, successful output of fully automatic machine translation (FAMT) has proven to be a reality only under very strictly defined conditions. In all remaining situations, different approaches need to be taken to optimize the efficiency of an MT system. At least, this should be the ideal if automatic translation wants to compete equally with the high quality offered by a professional human translator. The following are some options to improve MT effectiveness.

My masters program in Translation Studies at Dublin City University (Ireland) focused on MT dictionary building and updating combined with post-editing for improving the back-end results of the MT process. I realized that a profitable implementation and understanding of this technology means looking at both its reasonable potential and current limitations. MT is not perfect, but we must learn how to benefit from this imperfect tool that, in some situations, is justified by an economic necessity.

**Why MT Post-editing?**

Post-editing as a step or set of steps in an overall translation process is the term used for editing, modifying and/or correcting machine-translated texts. Although there are different types of post-editing serving different purposes, the distinctions among these levels and the tasks to be performed with regard to changes are not very clearly defined. This lack of clarity is the reason why sometimes full post-editing is performed, even though rapid or minimal post-editing might only be needed.

After an exhaustive review of existing articles and books on the topic, I quickly recognized that hardly anyone has taken time to carefully benchmark the speed of human translation as opposed to MT post-editing on any language pair. If and when such measurements have been made, the information is usually not publicly available. Most organizations or corporations implementing MT in combination with any type of post-editing do not necessarily make available any details or results of the process, the guidelines used or the training materials. Given that this is a significantly unexploited area of very practical and necessary research in the translation industry, I decided to focus on answering the key question: Is MT and full post-editing of raw MT output really faster than human translation from scratch?

Measuring the time of the different steps of the process was a main objective of my thesis research. Other information, such as the number and type of changes made, was also gathered. Money or cost savings as a further variable was not addressed in detail. Nonetheless, since time and money are directly related in this field, the decision to focus the study on the first factor implied indirect coverage of the second one. In addition, issues such as levels of post-editing, who should be doing post-editing, ideal skill sets, translation pricing schemes, the importance of building and updating dictionaries, and post-editing versus human translation and revision were also discussed. A chapter was also provided with a review of the MT software that was used, @promt professional English<>Spanish. A few suggestions on how to build or code entries into an MT dictionary, together with some guidelines on performing full post-editing, were also included. Since there was no objective way of measuring the quality achieved in the corresponding human and machine-assisted translations, two additional appendices were attached for further checks.

**A Challenging Domain**

The field of “marketing brochures” was chosen for conducting comparative statistics on human translation speed as a benchmark for carrying out MT post-editing measurements. Based on a study in which the author conducts a number of production mode experiments with French-English MT post-editing, the aim of my work was to add a set of English-Spanish statistics considered necessary for future comparative studies.

The most interesting point was testing MT efficiency for the translation of texts that were not highly repetitive technical and/or scientific manuals. It has been generally stated that MT systems are more appropriate for areas where the terminology is significantly and sufficiently repetitive, with a consistent writing style, and for a domain which one hopes is controlled in some manner. However, marketing brochures seem to be at the other pole since they are written neither in a straightforward manner nor with a

<table>
<thead>
<tr>
<th></th>
<th>Brochure 1</th>
<th>Brochure 2</th>
<th>Brochure 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language direction</td>
<td>English&gt;Spanish</td>
<td>English&gt;Spanish</td>
<td>English&gt;Spanish</td>
</tr>
<tr>
<td>Number of words</td>
<td>841</td>
<td>994</td>
<td>635</td>
</tr>
<tr>
<td>Human translation time</td>
<td>1 hour 43 minutes</td>
<td>3 hours 26 minutes</td>
<td>1 hour 35 minutes</td>
</tr>
<tr>
<td>MT Post-editing time</td>
<td>1 hour 1 minute</td>
<td>1 hour 29 minutes</td>
<td>41 minutes</td>
</tr>
<tr>
<td>Time savings</td>
<td>42 minutes</td>
<td>1 hour 57 minutes</td>
<td>54 minutes</td>
</tr>
</tbody>
</table>

Comparing results
different lines of business. All information gathered, as presented in the table “Text corpora,” refers only to usable words in electronic format. Besides, all Spanish brochures (47,715 words) were used as translation evaluation metric for the English-to-Spanish translated texts. The results of the post-editing and human translation comparative tests are shown in the table “Comparing results.”

A professional translator was involved in the translation of the first and second marketing brochures. The idea was to get independent information on how long a traditional human translation process takes given a translator who has not already become familiar with the source text through, for example, the MT post-editing process. This choice also allowed performing MT post-editing without being familiar with the source text through, for example, the human translation process.

Methodology

With the permission of the author of this methodology, I decided to apply, test and build upon it in modifying the following variables: language pair, direction and corpus of texts. The base methodology has already been implemented in international organizations and some corporate environments and thus demonstrates a credible process for real-life contexts, rather than only being a theoretical approach for translation research. See the diagram “Multistep process in MT post-editing.”

Tests and Results

Thanks to a collaborative effort conducted with different companies, I created a corpus based on marketing information across fairly neutral style. This document type, which has previously been atypical in the MT field, seemed to be a challenging area to investigate the true usefulness of MT systems.

Although the translator was quite efficient due to his familiarity with the terminology as well as the content that was related to his area of expertise (computational linguistics), the post-editing process was 42 minutes faster. In the second example, where the experienced translator needed more time to search for the meaning of unknown terms as well as for translating the source text, the value, efficiency and accuracy of the software were clearly revealed. His professional human abilities could not compete with the post-editing that was 1 hour 57 minutes faster.

I also conducted human translation tasks to compare with MT post-editing. The reason for this second choice was based on the notion that in industrial environments where MT is implemented, companies do not usually employ novice translators to perform MT post-editing. They mainly rely on experienced translators who are already familiar with the domain and texts that they are expected to translate. Even in this case where I was the only person conducting both processes and considering that AMT post-editing could have been performed faster if dealing with a previously (human) translated text, I considered not counting the time spent on reading the source text or dictionary research (a total of 19 minutes). The post-editing process was still 35 minutes faster.

In general, competing with an experienced translator or having previous familiarity with the text, MT in combination with full post-editing implies, at least, time savings of 42 minutes every 841 words or 35 minutes every 635 words. Taking into consideration previously published translation speed statistics and additional survey results conducted among translation organizations and independent translators, an average translator produces between 2,000 and 3,000 target language words per day. Thus, according to the data obtained in this study, time saved when conducting MT post-editing will range from approximately two to three hours: more precisely 1 hour 50 minutes to 2 hours 45 minutes (for the same number of words).

Based on these statistics and depending on specific needs and translation demands, any corporate manager can proceed with calculating the return on investment. My question is: If a similar high-quality published document can be achieved in less time using a semi-automated process, why not opt for the faster process?

Conclusion

Some conditions and constraints must be considered to make the post-editing of machine-translated texts worthwhile and effective. If these factors are observed, the efficiency of an automated translation system can be better understood and implemented even for translating marketing brochures. This demonstrates the efficiency of MT systems for such texts that are other than highly repetitive technical manuals and/or scientific documents.

It appears that some of the lack of interest in MT among multinational corporations, translation agencies and freelance translators is mainly due to a lack of information about how to conduct MT post-editing. This work has aimed at offering an in-depth set of statistical information that has not yet been provided to the MT user community during the past 50 years of developing and implementing MT systems, nor the past 30 years of investigating post-editing. This study is another step toward providing fundamental comparative statistics for further research and implementation case studies.
The Case for Proofing 100% Matches

Hélène Wimmerlin

The emergence of translation memory (TM) technology brought with it the possibility to increase consistency, yet also to reduce turnaround times and costs. Whenever the same or a similar phrase recurs, the TM automatically suggests the memorized translation as a possible match. As a result, the same sentence never needs to be translated twice. Translation agencies started embracing the technology and explained it to their clients as a way to improve the quality of translations while allowing the agencies to offer cheaper rates.

The technology was fairly new, and wary users weren’t necessarily knowledgeable enough to know its limitations — preferring, therefore, to err on the safe side and have human translators verify all the work done by the computer. They would in particular review the 100% matches. These are also called exact matches, and they are segments in the document being translated which are absolutely identical in language and format to a segment in the memory. The software is, therefore, able to plug in the translation automatically in the document.

The more widespread use of TM tools, the harder economic times and the increasing understanding of the available technology called for new ways of saving on costs. Not reviewing the 100% matches seems to be the latest trend as the usefulness of such review is considered debatable. Skipping the review of the exact matches has two major advantages: it saves time during the course of the project, and it allows yet another cost reduction because there is one less task to perform. This approach takes for granted the good results from TM systems. If the segments in the TM are well translated, then why worry about reusing those segments without validating them?

A reason to worry would be that a TM system is just a piece of software that doesn’t have the insight and understanding a human being might have. Let us remember the purpose of TM tools: they are there to assist translators — not to replace them. No TM software claims that the human validation is unnecessary. The match degree only indicates a match between segments taken out of context. It is that lack of context which is at stake here. Two identical segments might have different meanings or different grammar in different contexts.

First of all, a segment is not necessarily a sentence. A segment is a unit of translation. Not all text, however, is written in a sentence form. A segment can be part of a sentence or a phrase. Segments are defined by the segmentation rules of the TM software one uses. The end of segment markers can be periods, paragraph marks, exclamation points, question marks and so on.

A period is, however, not always the end of a sentence (for example, in abbreviations). Some TM software breaks the expression Max. height into two segments because of the period. If you translate this in French, it becomes Hauteur max. Say we now need to translate Max. length. The TM will bring the wrong translation even though the Max. segment will appear as a 100% match.

The way around this is to join the segments back together before validating the translation and entering it in the TM. However, this isn’t always possible due to occasional software limitations. In addition, not all translators are aware of the implications of joining, and some might skip this step.

Matches also need to be reviewed in context to make sure they fit the style of the text around them. Forcing the translator to look at the text before and after the segment helps him or her maintain the same style. For example, in the case of instructions, many languages can use either the imperative or infinitive form of a verb. If the translator chooses one form to translate a document and then the exact match is in a different form, he or she needs to make sure this is caught so that one doesn’t end up with something like this:

<table>
<thead>
<tr>
<th>No match</th>
<th>Open the file.</th>
<th>Ouvrez le fichier. (imperative form)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% match</td>
<td>Select the text.</td>
<td>Sélectionner le texte. (infinitive form)</td>
</tr>
</tbody>
</table>

The grammar of a 100% match segment could also need some adjustments. A sentence that was correct might be wrong in another context.

The pipe is rusty, = Le tuyau est rouillé. It needs to be replaced, = Il faut le remplacer. (masculine form)

Here is the second segment in another paragraph:

The handle is broken. = La poignée est cassée. It needs to be replaced, = Il faut la remplacer. (feminine form)

The item that it refers to has an implication about its gender in many languages other than English. The translator should pay special attention to pronouns and variable words whose antecedents are in previous segments.

Any incomplete sentence should also be carefully reviewed to make sure it fits the style and grammar around it. Incomplete sentences in a bullet point list are especially prone to needing fixes as they depend upon other segments. Catalogs are in particular full of incomplete sentences and require more attention on the part of the translator.

Another point of importance is that no translation is ever perfect. A fresh look at previous translations might allow for catching errors. Based on clients’ procedures, it is not always an easy process to change terminology in a 100% match, but translation is about refining the use of words.

The cases when errors are found are limited, but the errors are existent. 100% matches aren’t to be 100% trusted. Not reviewing the exact matches is taking a chance, hoping that erroneous translations such as I have described here have not occurred.

A different approach should be to review the 100% matches but not in the same way as the rest of the text. Every word in an exact match does not need to be put in question the same way that it would be when proofreading the document. Things to look for are variable words and their antecedents; incomplete or broken sentences; context; and style.

In addition, the grammar of a 100% match segment could also need some adjustments. A sentence that was correct might be wrong in another context. The pipe is rusty, = Le tuyau est rouillé. It needs to be replaced, = Il faut le remplacer. (masculine form)

Here is the second segment in another paragraph:

The handle is broken. = La poignée est cassée. It needs to be replaced, = Il faut la remplacer. (feminine form)

The item that it refers to has an implication about its gender in many languages other than English. The translator should pay special attention to pronouns and variable words whose antecedents are in previous segments.

Any incomplete sentence should also be carefully reviewed to make sure it fits the style and grammar around it. Incomplete sentences in a bullet point list are especially prone to needing fixes as they depend upon other segments. Catalogs are in particular full of incomplete sentences and require more attention on the part of the translator.

Another point of importance is that no translation is ever perfect. A fresh look at previous translations might allow for catching errors. Based on clients’ procedures, it is not always an easy process to change terminology in a 100% match, but translation is about refining the use of words.

The cases when errors are found are limited, but the errors are existent. 100% matches aren’t to be 100% trusted. Not reviewing the exact matches is taking a chance, hoping that erroneous translations such as I have described here have not occurred.

A different approach should be to review the 100% matches but not in the same way as the rest of the text. Every word in an exact match does not need to be put in question the same way that it would be when proofreading the document. Things to look for are variable words and their antecedents; incomplete or broken sentences; context; and style. £2
Translation Speed vs. Content Management

Jeff Allen

An organization informed me some months ago that it was conducting a machine translation (MT) pilot study. When I asked what their human translation (HT) speed statistics were, they responded that their translation project teams can handle X thousands of words translated per year. I was shocked by this reply because annual content management is simply a summary statement from the project management perspective of many different related tasks. Elements include reading and comprehension time, translation throughput speed, background research, terminology database management, sentence alignment, customer relations management, multiple review cycles and so on over a long period of time that can easily also include many activities unrelated to these translation tasks.

A tool evaluation/trial conducted during a several-week period should be based on measurements of comparable elements and factors. Comparing a one-day or one-week set of translation statistics against yearly translation content management figures is a risky way of trying to show that MT is more effective than only HT processes. I fear that many organizations fall into this same approach of conducting a comparative pilot study of HT vs. MT.

From April 2003 through October 2003, I distributed an informal survey to many organizations and international organizations fall into this same study of HT vs. MT.

I have two concerns from all of these data. First, how can many multinational corporations and government units and international organizations establish deadlines without some indication of their volume production? Second, how can any large organization or translation provider, those who are most likely to consider implementing an advanced language technology solution, proceed with translation tool evaluation and implementation actions without first having some low-to-high range statistics on volume produced in their current translation cycles?

I believe that such measurements are critical for translation task and resource forecasting and for providing an honest evaluation of the translation tools to be implemented. To demonstrate this, I will draw an analogy from a part of our daily professional lives that can be measured in terms of word level versus overall content management: typing speed. We can measure typing speed in terms of words per minute or in terms of pages of material actually typed (produced) per year. Let me provide a historical view to show some factors and variables to consider.

Twenty years ago, I took three full semesters (1.5 years) of typing courses at junior high school (for children ages 12-15 years old). During the first semester of the course, I achieved a typing speed of 32 words per minute (wpm) with 0 errors on a manual typewriter. During the second and third semesters, I reached a typing speed of 72 wpm with 3 errors on an electric typewriter. All of these statistics were based on printed documents during typing tests. It is also important to note that these tests were also based on short-term sprint (or burst) typing for less than 5 minutes, instead of overall general typing speed for an entire day. I will also add that the current world record for one minute of burst rate typing is 216 wpm, and the world record for sustained typing during a period of 50 minutes is 170 wpm (37,750 words). Also important to note is the fact that such typing is based on reproducing existing texts; it does not include any creative “cognitive” activity, which is part of the translation task.

During my first year of college/university studies, I bought an electronic typewriter that displayed an entire line of memory on an LCD screen before it printed to the page. This is where the typing speed and typing accuracy ratio started to become affected. I had the chance to go back and clean up mistakes for an entire line of text before they printed to the page. So, initial overall typing speed may have increased because I knew that there was the possibility of checking the line before printing it to the page, or it may have decreased because doing the additional checking on the miniature screen was an additional constraint. In either case, the overall typing speed (based on the typing and pre-print editing time) might have decreased, yet accuracy increased overall for volumes of text that were much longer than what can be typed during one minute of time.

Then, during my last year of undergraduate college studies, I bought an Apple II computer. This tool had certain effects on typing speed. My personal typing speed increased significantly, yet my typing error rate also increased. The reason for this was simply that I could review the complete text before printing or could use the spellchecker afterward to semi-automatically clean up typing errors. With the advent of the computer, too, the issue of precision and accuracy was
All of the raw figures were rolled up or drilled down to fit along a standardized axis of words per hour, words per day or pages per day. Weekly vs. daily vs. hourly figures: weekly statistics were converted to daily figures and daily figures to hourly values based on an 8-hour work day. Words per page: a total of 178.5 words per page was calculated based on information found in existing statistics combined with respondent replies to the survey. The set of replies on page count contained number of characters per line multiplied by the number of lines per page. This was then divided by 6.8 words per line based on previously conducted surveys. These values were standardized and aggregated to provide a single average value of 178.5 words per page. Rounding off time periods: In cases where time periods were other than hourly or daily (such as a 45- or 50-minute period), a recalculation was made to standardize the value to an hourly basis.

From these multiple factors, we see the following results over a 20-year period.

- Overall typing speed in English has “decreased” by 15%.
- Ability to type 50 wpm in another language (French) at end of total period whereas the language skills in the second language were nil at the beginning of the period.
- Typing speed has decreased because the typing activity no longer focuses on reproducing printed text, but rather now on simultaneously creating new information.
- Word content management has “increased” exponentially because of leveraging thousands of text elements as a memory database in order to avoid retyping previously created content.

So now let's consider these statistics in a real production environment scenario. If I am faced with the need to provide an estimate of how much time it would take to write up a project proposal or respond to a very long request for information (RFI) or request for quote (RFQ), I would be able to make the following statements based on known statistics:

If I have a database of existing material and I am a subject matter expert in the field, then I can produce X amount of texts in Y amount of time.

If I have a database of texts, but I do not know the field, then I need to add at least X%
of time to my statistics above to take into consideration the learning curve impact.

If I know the field, but no texts are available on the topic, then I can base the proposal preparation time on 50-60 wpm typing rate, plus factoring in the time for conducting Internet research and some cognitive effort to create new content.

If I do not know the field well and no texts are available on the topic, then I must at minimum base the preparation time on 50-60 wpm typing rate, plus intensive Internet research, plus significant cognitive effort.

In any case, it is virtually impossible to consider measurements beyond the world record of 150 wpm of sustained typing (for 50 minutes) or the world record of 216 wpm of burst rate typing, or faster than the world record of 12-14 notes per second of piano trills.

This shows that typing speed, although a very basic function which can be separated from the volume of material to be produced over a period of time, can clearly be used to determine resource allocation. Such a benchmarking measure is necessary when having to deal with requests for which existing texts or expert knowledge of the field might be lacking.

These statistics allow us to make the following time-related statements:

Basic typing speed is A.

When texts are not available and knowledge is low, then production speed decreases to B.

When texts are available and knowledge is low/medium, then production speed might only decrease to C.

When texts are not available and knowledge is medium/high, then production speed can increase up to D.

When texts are available and knowledge is medium/high, then production speed will likely increase significantly up to E.

A similar analogy can be made with the figures necessary to justify the purchase of a large printer or photocopier for business activities. One needs to gather individual statistics on the range of documents printed, how many pages per document, the frequency of (re)publication of the documents and so on. I will not go into detail, but I think many of us have been faced with such situations. From this information and other factors, it is possible to determine the cost of printing per page and decide whether or not the investment in the technology is worth the purchase.

The basic lack of any real benchmark and progress improvement measurements within the translation process (using CAT tools versus not using them) can be problematic. In contexts where the tools show good productivity because the conditions are favorable to text leveraging with TM or good MT output because of already existing terminology databases, then the productivity histograms and pie charts will look wonderful. Over time, the charts will look better and better, and it is possible to show the ability to handle 20%, 30%, 50% more words/pages per year than previous years. Yet, when a new (sub)domain or document type is introduced into the environment, which requires either significant upfront terminology work, full translation database build-up to even have usable TM leveraging material and/or text alignment work, then the previously used pie charts cannot be used as valid benchmarking figures. This leads to the possible risk of significantly underestimating the translation resource workload.

I believe that the same set of general measurements and statistics that can be taken from typing speed should be used for authoring/translation activities by separating the basic tasks from overall content management and taking into consideration the various other factors such as knowledge/expertise. Otherwise, language technology implementation decisions based on unequal measurements could have drastic results if the overall translation cycle is significantly crippled due to a poor evaluation method. I have conducted authoring and translation statistics benchmarking in the past, and these data led to improved use of the language technology tools that were implemented. Additional examples of such translation benchmarking projects will appear in the coming year.

We must keep in mind that language technologies have their constraints and limits, and it is thus important to know the true baseline in order to measure improved and lowered productivity at both task-specific and overall levels. This can help us justify the increase or decrease of productivity efforts at different levels of a given context rather than simply saying that the overall productivity has gone up or down. This is, therefore, a call for all translation and project managers to evaluate whether or not they really know their translation volume productivity statistics.

### Resources

| Association for Machine Translation in the Americas (AMTA), European Association for Machine Translation (EAMT) and Asian-Pacific Association for Machine Translation (AAMT) are regional associations of the International Association for Machine Translation (IAMT) | www.amtaweb.org | www.eamt.org | www.aamt.info |
| The IAMT sponsors the MT Summit | www.amtaweb.org/summit |
| Human Language Technologies focuses on European research and applications. | www.hitc.org |
| MultiLingual Computing, Inc. on-line resources and annual Resource Directory/Index. | www.multilingual.com |
| Translator’s Home Companion | www.lai.com/thc.html |
| University of Manchester Institute of Science and Technology (UMIST) Center for Computational Linguistics | www.ccl.umist.ac.uk/dle/index.html |
| Jeff Allen’s MT pages include links to an extensive list of publications in the field of machine translation. | www.geocities.com/mtpostediting and www.geocities.com/langresourcesallen |
| Word2Word Language Resources | www.word2word.com/m.html |
| Translation organizations such as American Translators Association (ATA) and Institute for Translation and Interpreting (ITI) | www.atanet.org | www.lti.org.uk |
| Individual companies such as Applications Technology, Inc. (AppTek); Ectaco, Inc.; SDL International; SYSTRAN Software, Inc.; WH&P, Inc. | | | |
A Translator Protests Poor Use of Machine Translation

Carlo Capussotti

Is it possible to integrate machine translation with the translator’s work? Maybe, but it mostly depends on the quality of the MT program. Some products give results that are at least understandable, whereas others simply dish out incomprehensible words. And even the more sophisticated programs can be tricked by ambiguity and make translation errors.

Let’s look at a classic example. The Russian company XYZ decides to translate the advertising material (written in Russian) for its cars into French, Italian, German, and Spanish. Seeing that translating from English is cheaper and more practical than translating from Russian, as a first move XYZ decides to translate its advertising brochures into English so that the translation into the other languages can be based on the English text.

The marketing department decides that having the text translated into English by a translator would be too expensive and that one of the leading products in the MT sector guarantees perfect translations for just a bit more than $300. The decision is made, and the text is machine translated.

The “English” text is then sent, without any further checking, to localization agencies in the various countries for translation. Agencies send the text to their freelance translators, but the text makes hardly any sense. Here, for example, is a Russian sentence that was taken from the Web site of a car maker and translated using the demo version of an MT software: “We present on the North Western market of Russia the production of the vehicle spare parts leading producers. In the years of chuushchestvovaniya our clients became the station of maintenance, motor transport establishment, the stores, the wholesale firms.”

If the translator knows Russian, he or she could try to deduce the meaning from the original text if it was provided; otherwise, there is no other way than guessing.

Here is another case. A well-known British localization company (Agency A). Agency A subcontracted the other languages to smaller agencies (B) in the different countries. Italian was one of the languages, and the example refers to this language in particular.

The project was large and had to be translated using a CAT program. Because this was an updating process, however, translators were asked upon Agency A’s request to ignore all 100% matches and deal only with new or partially new segments.

As soon as the project started, project managers (PM) of an Agency B were informed by translators that the pre-translated material (100% match) was almost entirely unusable due to poor quality. It is true that 100% matches had to be ignored, but this material was going to become part of partially translated sentences. All the translators and the PMS agreed that the text had not been translated by a “human” but rather by an “inorganic entity.” Agency B and Agency A were notified, and the project was temporarily suspended.

While neither confirming nor denying the use of an MT program when pre-translating, Agency A requested that the 100% matches also be checked during the translation — without, however, significantly changing the delivery date. Since the pre-translated material was completely unusable, Agency A was basically asking for a total retranslation, giving a deadline that would normally be enough just for proofreading the text.

The work was done in haste, and Agency A was not satisfied with the result. Agency B organized a conference call with Agency A, but the latter prohibited PM’s from participating on the grounds that the problem was not technical but commercial.

Finally, Agency B deducted 10% from the agreed amount, due to alleged “poor quality,” throwing the entire blame on the translators and ignoring the initial problem, that is, the machine translation.

MTs are not a shortcut to easy money. MT may be of help to grasp the general meaning of a text, but it must always be combined with the work of a professional and must not be seen as a total alternative to the work of the translator.

MT Requires More Proofreading

If MTs are used correctly, costs are higher than with “traditional” translations because, although the steps to go through are the same, the time needed for proofreading is longer.

The MT program does not produce a translation but just words. Basically, it is a sophisticated dictionary that transforms the words from one language into another.
Sometimes making gross errors. It is the task of the professional to correct, wherever possible, these errors and edit the text in those cases in which proofreading is impossible.

Comparing the two processes, how high will the final quality obtained from each process be? If the initial quality of translation (phase 1) is low, all of the following phases will be affected. The complete retranslation of the text in a later phase is unthinkable; therefore, the proofreader will have to try to make the machine-translated text acceptable. The final quality of the translation also obviously depends on proofreaders.

Let’s go back to the example of the car manufacturer. This company has a catalog of spare parts whose names remain more or less unchanged over time (right front disk brake, lambda probe and so on). Such a repetitive text can certainly be translated with a good MT program provided with an automobile glossary and programmed to recognize the terminology used by that specific company.

This does not mean that the final result should not be proofread by a proofreader to avoid translation errors. In this case, however, seeing that the text is not complex and consists only of simple words, the checking process will be very fast, and the final result will be acceptable.

So, the quality of the final product depends not only on the translator but also on the ability and competence of the proofreader.

**Proofreaders**

Who is the proofreader and what does he or she do exactly? Even many proofreaders are not sure about their role.

The proofreader can be an independent translator who has a good knowledge of the product or a co-worker, often from the client’s marketing department. His or her task is to make sure that no mistakes have slipped by in the previous phases of translation, DTP, formatting and so on; correct them if necessary; and give the go-ahead to the production phase.

The translated text is read from the user’s point of view; concepts that are difficult to understand are highlighted and improved; and mistakes that may have slipped by during the translation process as well as errors in the product’s specific terminology are corrected. It is important to avoid changing the legacy, especially when the terms are in the glossary. One must avoid replacing a word with one of its synonyms. If a word is correct, replacing it with a synonym is certainly not going to improve the quality of the manual.

If, for whatever reason, the decision is made to change the terminology or the style, the consistency of corrections is fundamental for the quality of the product.

While proofreading, the temptation to rewrite everything in one’s own personal style is extremely strong. This is not the purpose of proofreading, and there are reasons why this is so. In general, the proofreader does not know whether the text is part of a larger project, and therefore by changing the style, inconsistencies may arise in the rest of the project. In addition, even if the terminology is not exactly the one the proofreader would have chosen for the document, it might be the terminology that the client has used for many years and is completely happy with. When one is about to make a correction, it is a good idea to ask whether it is really necessary. If the answer is no, then the correction should not be made.

If the translation has been done correctly, the number of corrections made by the proofreader should be minimal. I do not think that it is the proofreader’s task to try to save a badly done translation, especially when the proofreader is not a linguist. Often the proofreader is not provided with the original text, thereby making it difficult to correct a translation where, for example, there are interpretation errors. It is better to send the text back to the translator asking for a revision of the translation.

Unfortunately, the relationship between translators and proofreaders is not always idyllic. Both translators and proofreaders are human beings and, as such, make mistakes.

To go back to the MT issue, proofreaders may find themselves having to correct an MT text. The proofreader must always be a translator of the languages concerned. If the translation is from Japanese into Turkish, a Japanese-Turkish translator is necessary. It is unthinkable to entrust the proofreading of a text almost certainly containing large translation errors to someone who is not a professional in the two languages. In this case the proofreader will be the sole judge of the situation as, for obvious reasons, there is no feedback from the translator.

In conclusion, I feel that there are areas where MT is both usable and useful. These areas are restricted to those cases where the translation does not need to be of high quality and only a general idea of what is written in the text is needed or where simple words need translating rather than a complex text.

The translated text must always be submitted to a competent translator/proofreader to transform a product from a very sophisticated dictionary into a translation. Ω
The Electric Word and the Semantic Web

Jaap van der Meer

Some diehards in the translation industry may remember the magazine Electric Word, preceded by Language Technology and later reincarnated as Wired under the inspiring leadership of Louis Rossetto. The metaphorical title Electric Word was intended to show wordworkers that technology was soon to move beyond merely the processing of words. Nobody imagined at the time, however, that the word would ever get a life of its own and that the word would indeed one day be electrified and start interacting with other words, applications and computers without human intervention. That would have been a horrifying vision for all the wordworkers in the world, translators and localizers included. Is it still, perhaps? And yet, that is what is about to happen right now.

Words have been dead “things” for as long as people worked with them. Words only get a meaning in the minds of people, never on paper or in computers. For people to communicate properly they need to agree on the meanings of words. This has never been a real problem in the history of human civilization because we have great skills for interpreting the meanings of words. We use all our senses and even our intuition to quickly build relationships around the words that are presented to us. This is how we can make a difference between capturing data and capturing criminals or saving files and saving money or edible cookies and cookies on a computer. We just know it, right?

Well, sometimes we do not. Technology development is just going too fast. Every year, every half-year, a new product release, a new PDA, a new OCR, a new wafer stepper, a new what? And technologies are blended. Cars become computers; mobile phones become digital cameras; computers become home entertainment centers. What’s more, technology is crossing borders faster than ever before. So, how do you tell the difference among a GSM, a mobile, a cell phone or a handy? Or are they all the same thing? Why do we not simply standardize terms for new things? This would make the work of translators a lot easier and put an end to the confusion of end users. Humans, however, are not all that rational. It is in our nature to conform ourselves as much as to differentiate ourselves or rather to conform ourselves to other cultural categories.

Information keeps piling up in our companies as more and more colleagues, suppliers and customers write documents, e-mails and forms that become part of our corporate knowledge base. But what knowledge do we really have? Meanings get lost in this highly dynamic environment. Explosive growth of information, blending of technologies, rapid innovations and human nature simply resist conforming to standard words. For a long time we turned to the professional wordworkers to guide us through the jungle, to help us differentiate a wafer stepper from a toaster in thirty languages. But this task simply becomes too daunting for humans alone. Glossaries could not help; they are tracking instead of prescribing which terms we use.

The rapid proliferation of HTML and XML as standards for structuring documents allows searches on words and strings of words. But these techniques still look at words as dead digits. This is why a search on wafer stepper returns thousands of results with perhaps only five relevant to the question asked. The pile of information is still too big for humans to digest.

The solution to this problem is simple and powerful: every word must get its own digital passport or unique identifier. This passport should contain information on the relationships that words may have and perhaps other information such as domain attributes and properties. Words would no longer travel as dead digits but would carry their own meanings and be able to freely establish relationships and create more sense.

This is just one step away from the great new vision for the World Wide Web. The Web must become a Semantic Web, as expressed by Tim Berners-Lee and his colleagues at the World Wide Web Consortium (W3C). On the Semantic Web every word will be tagged with information: parentOf, childOf, partOf, has Location, worksFor, hasSubjectOf, dependsOn and so on. Unlike the practice in conventional dictionaries and glossaries, this information must be described in machine-readable form so that computer applications can start exchanging information without human intervention.

The W3C, together with many universities and corporations around the world, has started building the key components to make the vision of the Semantic Web come true. Basic concepts are to be defined in ontologies. People may use different words, but the semantic tagging will invoke the relationships and refer to the basic concept as it is defined in the ontology. Ontologies are already developed for several industry domains such as the financial and medical industries.

Once these key components are in place, we can imagine Web portals where semantic identifiers lead users automatically to the most relevant information. The effectiveness of searches can be further improved by matching user profiles with semantic identifiers. We can imagine that computer applications will "talk" to each other to schedule meetings, to analyze data, to compose reports and translate documentation, all without human intervention.

Linguistic intelligent searches and Semantic Web projects all originate outside the traditional language service industry. Knowledge managers in many companies start to build taxonomies and thesauri to help make sense out of the vast amounts of information on the corporate Intranets and knowledge bases. When the time comes, they are ready to connect to the Semantic Web.

And what do the wordworkers do? Are they unaware of this grand linguistic opera that is undertaken in the world of technology? Or are they perhaps paralyzed by this horrifying reality of the "electric word"? After all, it was they, the writers, who had always mastered the meaning of words.

It may be ignorance or it may be fear; the truth is that the translation industry is hardly involved in these ambitions for corporate taxonomies, industry ontologies and the Semantic Web. And yet, there have never been more exciting times for language professionals. Imagine how “electric words” can fine-tune matches in translation technology. Semantic relationships and domain attributes stored in the digital passport of every word ensure that the accuracy of MT and search will improve dramatically. Writers and translators should get over their fear and join forces with knowledge managers to empower the word. Their linguistic skills are indispensable.
This supplement introduces you to the magazine *MultiLingual Computing & Technology*. Published nine times a year, filled with news, technical developments and language information, it is widely recognized as a useful and informative publication for people who are interested in the role of language, technology and translation in our twenty-first-century world.

### Translation

How are translation tools changing the art and science of communicating ideas and information between speakers of different languages? Translators are vital to the development of international and localized software. Those who specialize in technical documents, such as manuals for computer hardware and software, industrial equipment and medical products, use sophisticated tools along with professional expertise to translate complex text clearly and precisely. Translators and people who use translation services track new developments through articles and news items in *MultiLingual Computing & Technology*.

### Language Technology

From multiple keyboard layouts and input methods to Unicode-enabled operating systems, language-specific encodings, systems that recognize your handwriting or your speech in any language — language technology is changing day by day. And this technology is also changing the way in which people communicate on a personal level; changing the requirements for international software; and changing how business is done all over the world.

*MultiLingual Computing & Technology* is your source for the best information and insight into these developments and how they will affect you and your business.

### Global Web

Every Web site is a global Web site, and even a site designed for one country may require several languages to be effective. Experienced Web professionals explain how to create a site that works for users everywhere, how to attract those users to your site and how to keep it current. Whether you use the Internet and World Wide Web for e-mail, for purchasing services, for promoting your business or for conducting fully international e-commerce, you’ll benefit from the information and ideas in each issue of *MultiLingual Computing & Technology*.

### Managing Content

How do you track all the words and the changes that occur in a multilingual Web site? How do you know who’s doing what and where? How do you respond to customers and vendors in a prompt manner and in their own languages? The growing and changing field of content management and global management systems (CMS and GMS), customer relations management (CRM) and other management disciplines is increasingly important as systems become more complex. Leaders in the development of these systems explain how they work and how they work together.

### Internationalization

Making software ready for the international market requires more than just a good idea. How does an international developer prepare a product for multiple locales? Will the pictures and colors you select for a user interface in France be suitable for users in Brazil? Elements such as date and currency formats sound like simple components, but developers who ignore the many international variants find that their products may be unusable. You’ll find sound ideas and practical help in every issue.

### Localization

How can you make your product look and feel as if it were built in another country for users of that language and culture? How do you choose a localization service vendor? Developers and localizers offer their ideas and relate their experiences with practical advice that will save you time and money in your localization projects.

### And There’s Much More

Authors with in-depth knowledge summarize changes in the language industry and explain its financial side, describe the challenges of computing in various languages, explain and update encoding schemes and evaluate software and systems. Other articles focus on particular countries or regions; translation and localization training programs; the uses of language technology in specific industries — a wide array of current topics from the world of multilingual computing.

Nine times a year, readers of *MultiLingual Computing & Technology* explore language technology and its applications, project management, basic elements and advanced ideas with the people and companies who are building the future.

---

**Subscription Offer**

To subscribe, use our secure on-line form at [www.multilingual.com/subscribe](http://www.multilingual.com/subscribe).

Be sure to enter this on-line registration code: `sup62`