

Aslib 95
Translating and the Computer
London, November, 1995

"Some practical suggestions for implementing machine translation"
John Hatley

Virtually any speech about machine translation with a time limit has to begin somewhere in the middle. Even a definition of the term can take a while but "beginning in the middle" is easier considering the current listeners. Nevertheless, I shall attempt to begin as close to the beginning as possible.

Machine translation is itself a sub-category of the much larger field of natural language processing (NLP), which includes, in a sense, any computer program which allows a user (expert or non-expert) to do *something* with natural language. Depending on how broad or narrow a definition I choose for machine translation, I can include or exclude any number of "natural language processing software" and thus influence the rest of the discussion.

So: for a working definition of machine translation let us call it **computer software for the automatic (*not* interactive) translation of text from one natural language to another, i.e. without human intervention in the actual translation process**. This definition excludes computer-aided translation systems, which are also often called interactive, i.e. require some amount of human intervention throughout the translation process, but it is not in any way intended to be a reflection of the quality or value of computer-aided translation. It merely makes it easier for me to talk about machine translation.

Having defined machine translation, there remains one general term which often requires clarification: **language pair**. This is the sum of one **source language**, the language from which the software translates, and one **target language**, the language into which the software translates. It is *not* "a pair of languages" as we often leave people believing, for which reason I prefer to use the somewhat clumsy term "**direction of translation**". We are now ready to make a few suggestions how to implement machine translation software. We must first pose a few questions, but the questions themselves represent a first step towards a possible implementation. Although I have given some thought to the order in which I ask the following questions, apart from the very first one, the order is more or less arbitrary. The first question is:

"Does my company *translate*?" Whereas it seems a bit silly even to ask the question, it is surprisingly often the case - particularly within larger companies, that there is little awareness of translation activities. It may help you to find an answer to this by asking, "Do my company's products or services require documentation? Does my company export its products or services to countries whose language is different from the language in which the documentation is written?" You must also ask the question concerning your company's direction of translation at the very beginning. If your company has a critical and exclusive need for Greek to Norwegian translation, you will soon find that there is no machine translation software for that combination of languages.

If the answer(s) to this first (group of) question(s), is "yes", then we can proceed. Finding out whether it makes sense to evaluate MT can take a considerable amount of time and effort, and we are only at the beginning.

Next come the *how much* questions: **"How much do we translate?"** How many words, lines, pages, documents, books etc. does my company translate per day, per week, per month, per year? And, **"How much does it cost?"** The responses to these two questions often surprise MT vendors - companies very often have only a very vague idea or no idea at all how much they translate, perhaps more of an idea how much it costs - although we really should not be surprised. Translation is still very often seen (if indeed it is *seen* at all) as a necessary evil, although the awareness of its importance is increasing. There is also a phenomenon I call "hidden translations", which brings us to the next question.

"Where are the translations done?" Is the translation activity centralized or decentralized? Does my company have a translation department? How many translators work in that department? Where is it? This can be in a variety of different places from *central administration* through *public relations, marketing and development* to *documentation*. Finding out where translation is done is especially important to the MT-vendor, but it should also help the potential buyer to evaluate the importance of translation within his or her company and, not least, to find out whether there is a budget, how big that budget is, who is responsible for it (whose approval must be obtained) and who else should be involved in the discussions etc. As concerns "hidden translations" are there any other departments (apart from the translation department) in which translations are being done?

"Is an *external translation service agency* (also) used?" Are translations done externally *in addition to* internal translation? If the answer here is "yes", then it will be helpful to determine what the reasons are.

Perhaps German-English, English-German translation work is done by the company's internal department and the services of a translation bureau are only employed for translations into and out of Japanese. It may also be that "high volume" translations are done internally and the "little extras" externally (or vice versa). Perhaps external translation is either faster or less expensive (or both) than internal translation. Or perhaps external translation is merely "more controllable" than internal translation.

If translations are done externally *instead of* internally and if the reason for this is the lack of translators within the company, this may present a serious problem for a possible implementation of an MT system, considering that an MT system must be operated by a translator (or at least a bi-lingual person). It is not impossible to make use of an MT system in a company which has no translators; a number of critical conditions must be met, however, for that use to be successful. This applies equally to a "mixed use" of machine translation: it is quite feasible for a company both to install translation software internally and continue to use the services of a translation agency which is *also* using *the same* machine translation software - in the same way in which many companies work today without MT.

"Are we currently using any *translation tools* (software)?" Especially as concerns the many very good terminology data bases, we must consider whether the terminology in that data base can be transferred *easily* to an MT system. This issue is equally important whether the plan is to *replace* the terminology data base with an MT system or to operate the two systems *in parallel*. The same applies, of course, to the current use either of interactive translation aids or translation memory systems. Both interactive translation and translation memory software can be extremely useful. Keep in mind, however, that neither of these tools is machine translation by our definition above.

Finally, **"What do we translate?"** I say finally not because there are no further questions to ask after this one, but to get round to some concrete suggestions more quickly. Any machine translation system (by our definition) should be able to translate any text I submit to it (after all, it's only a machine). However, MT systems produce at the very best a raw translation; the quality may be quite acceptable or even very good, but it remains a raw translation. This means that whatever an MT system translates and however good it is, it will have to be post-edited (corrected) by a human translator when the machine is finished. The amount of post-editing necessary varies greatly from one text (or type of text) to another. In short: MT is (still) best suited for the translation of technical documentation, for example product descriptions, users' manuals and installation and repair instructions. This does *not* mean that MT is not suited for other types of text. What it does mean is that

the machine translation of other types of text will require both more terminology work before translation and more post-editing after translation.

Machine translation software is *not* an off-the-shelf product which can be installed, switched on and used immediately to produce excellent translations. Its successful implementation requires a considerable amount of preparation and a willingness to make a continuous investment of time and energy.

Now for some concrete suggestions:

For a start: your company probably does not have a "translation problem", or at least it may not perceive it as a problem. Indeed, it is probably not the problem. Assuming this is a mistake we MT vendors often make. A company which manufactures video recorders may have a problem establishing a market presence abroad. But it may not see the contribution that translation could make towards a solution to that problem. It will help in the missionary phase of evaluation to remember this fact. In a very real sense, the company *has* no translation problem. Therefore, neither the translation staff nor the MT vendor should spend much time trying to awaken an awareness of a problem which perceptually does not exist.

Assuming we have done our homework and established that we have the need, the personnel, the volume, the direction of translation and types of text suitable - and, not least, the budget - for a serious evaluation of MT, we can begin to take some concrete steps.

1. Make a **realistic assessment of what you (or your company) expect** an MT system to do and whether any MT system can meet those expectations. An MT system *cannot* replace a human translator; it *can* be used by non-translators, but only if several crucial requirements are met first. No matter how large the MT system's dictionary is, it cannot "know" your terminology. This will have to be collected or compiled (if it has not already been compiled), in machine-readable form, and put into the MT system's dictionary. Ideally this should take place before installation of the translation software. If the user is expected to be able to enter terminology will have to freshen up his or her knowledge of grammar; it can translate any text you present to it *but* there are suitable and less-suitable types of text, which will require widely varying amounts of post-editing; it will *not* be able to translate into and out of every language; and it will not be able to run on *any* hardware. It *is* compatible with a wide variety of different word processing and electronic publishing software. But remember, the only hardware and software it *must* be compatible with is **yours**. Do not discount entirely

the necessity of purchasing a piece of hardware with a label which is different from the rest of your equipment. Such a necessity will, however, make the implementation of MT software more expensive than just the cost of the software. It is possible to make virtually any piece of hardware communicate with any other piece of hardware. It may be a little complex, but considerably easier than making two human beings communicate with each other.

What else *can* MT do?

It can translate extremely fast, so fast in fact that speed soon becomes irrelevant. Logos translates roughly 15,000 words per hour. Based on figures published by the German Bundesverband der Dolmetscher und Übersetzer e.V. (BDU), this corresponds to between 20 and 60 times the average daily production of a human translator (human translation approx. 7 to 10 pages per day; machine translation between 30 and over 80 pages per hour or between 240 and 640 pages per eight-hour day). Obviously, if the MT system produces between 240 and 640 pages a day of worthless rubbish, it does not make sense to install it. If, however, MT can enable a human translator to produce twice as much translation in a normal eight-hour working day as was possible without it, or to produce the same amount in half the time, it becomes considerably easier to argue in favour of it. Twice the volume or half the time is being achieved by MT users, especially for the suitable texts described above. It is also becoming increasingly possible for traditionally less-suitable texts. A human translator may also "produce" as little as half the "daily average" for translations of more intellectually challenging texts. Thus, it can easily take twice the normal amount of time necessary to post-edit the machine translation of such texts and *still* save time.

MT will also use your terminology with nearly perfect consistency. If you call a German *Bildschirm* a *monitor* on page one of your English translation it will not miraculously become a *terminal* on page 50 and a *screen* on page 100. This phenomenon may be considered desirable for stylistic reasons but it can genuinely confuse the reader of your manual.

How do you translate now? A detailed analysis of the translation **process** in place today will produce extremely helpful insights into how best to integrate a machine translation product into that process. It will show you areas in which the process may need to be adapted to the MT software and also where the MT software may need to be adapted to the process

2. You must make a careful and **realistic evaluation of the willingness of your company's translation staff to work with an MT system**. Actively involve your translators in the decision-making process right from the very beginning. He or she is the one who is going to be using it once it is installed! There is no point in forcing an "MT opponent" to work with MT. And there is little point in ignoring the fact that MT will dramatically change the way in which the translator has been working until now. It is wrong to assume that anyone who speaks two or more languages is automatically a good translator (or in fact that a good translator must *speak fluently* the languages into and from which he/she translates). It is equally wrong to assume that every good translator will be a good post-editor of machine translations or that a good post-editor is also a good translator.

3. Determine whether it is desirable, possible (or even necessary), and at what cost, to **transfer existing terminology lists to the MT system** being investigated. It is not absolutely necessary to put all of your existing terminology into an MT system's dictionary before productive work can begin. For one thing, some of the terminology may be outdated and thus never again occur in a new source document. For another, such terminology transfer is another cost factor - in addition to the cost of the actual MT software - because it is a service which the MT vendor is ready, willing and able to provide, but typically not free-of-charge. However, buying an MT system which already "knows" your company's specific terminology would certainly shorten the running-in time (which can take a few months). By running-in time I mean the amount of time it takes you to "teach" your MT system your vocabulary in order to advance from *acceptable* raw translations to *good* or *very good* raw translations. If your existing terminology lists are in a computer data base, also give serious thought to the question whether you want to continue to maintain that data base after the MT software is installed. If the answer is yes, this will sooner or later raise the issue of "communicating" between the terminology data base and the MT system's dictionary *in both directions*. Whereas the MT vendor will probably also be willing to customize a piece of software necessary to *export* the MT terms *to* the external dictionary, this service is likely to be more expensive than it was to *import* the terminology *from* the external dictionary.

4. Make at least an initial attempt to **determine the suitability of your documents for MT**. The only way to find out whether a new car you are about to buy will get you from here to there as comfortably or as fast as you hope, is to get in and drive it. Similarly, the only real way to determine the suitability of your texts for MT (or the suitability of MT for your texts - there is a slight difference!), is to translate them with MT. Considering the size of the investment, it is advisable, as a minimum, to

ask for a *sample translation* of a *typical* text, which will at least give you an idea of the translation quality you can expect on "day one" of the installation. You might also consider the possibility of arranging for a *test installation* of the MT software for a pre-defined period of time.

Although it is risky in this context to make general statements, I recommend a minimum of three months. If it is properly prepared, such a test installation will give you the chance to evaluate translation quality and text suitability; it will allow you to estimate how long the running-in phase will be, i.e. how much terminology work will have to be done to achieve acceptable quality and even get you started building up that terminology; it will also give you the opportunity to estimate the savings in time and money your company will achieve when you enter a production phase *and* to identify some of the problems you *did not* expect.

5. Early on in your evaluation, you must get some pretty definite answers whether the MT software has a filter which can translate documents written with *your* word processing or electronic publishing software. If it has not got one, it is not the end of the world - programs which can convert a document written with one word processing software to another word processing software are available and generally quite good. The absence of a filter for your word-processing software may, however, be the beginning of a potential difficulty: any time you convert a document from one word-processing package to another, something gets lost. If the loss is not detected and compensated for, it *will have* serious consequences for the quality of the machine translation output. It will probably also have serious consequences for the *format* of the output and a direct, negative effect on the amount of time necessary for post-editing. Keep in mind, however, that machine translation is extremely fast. The computer can produce raw translations at such an incredibly fast rate, that you may still save time even if there is an inordinate amount of manual re-formatting necessary.

6. The hardware question is also important but may not necessarily be as crucial as the software issue. At this point in discussions with potential customers I often pose the question: "What is the real issue? Is your company's problem 'strict adherence to a restricted hardware environment' or 'finding a solution to a translation problem'?" This question is genuinely not intended to be as cheeky as it may appear. If your company has a policy that any and all software *must* run on an XYZ computer and the MT software vendor you are evaluating only runs on an ABC computer, then that MT software obviously does not adhere to your company's policy. On the other hand, it is possible to connect an ABC computer to an XYZ computer (or to an XYZ

computer network) and it should be noted that this is often (though not always) less complicated than converting one software to another.

7. If it is not already the case, you should strive for a closer **integration of the documentation and translation processes**. Encourage your translators to make constructive criticism of the language of the original documents. Encourage your technical writers to accept that constructive criticism and to make suggestions of their own. The positive results of doing this are multiple. A thorough evaluation of the linguistic quality of your company's original documents over a longer period of time and concrete suggestions for positive changes should lead to an improvement not only of the translations but also of the original documents. It is *very* important to point out that I am not talking about a general *simplification* of language (although it may lead to that to a certain extent) but rather a *clarification* or perhaps *disambiguation* of language. It has long been the case that the human translator very often *improves* a document while translating it. Or to put it another way, the translation of a user's manual or technical document is very often more understandable than the original. How does the human translator do it? If I cannot understand the original of document I have to translate, I usually ask the author what was *meant*. What I then actually translate in such a case is some derivative of the author's explanation of what he meant. This phenomenon often prompts the question: "If that's what you meant, why didn't you write it?" Allow me to emphasize at this juncture that **an MT system can only translate what the author *meant* if that's what the author *wrote***. On the other hand, allow me to make an appeal to the translators: "Don't overdo it." For many types of document, the time it takes to complete a translation has already become far more important than the subjective quality of the language. A few years ago I was witness to a discussion between three translators about the appropriate translation of one *word*. The discussion lasted half a day. The word in question was the heading of a 35-page document. It should not be surprising that the department requesting it waited more than three months for the translation. If we go just one step further: taking into account the monthly salaries of those three translators and the average number of working days in a month, the translation of that single word cost their company DM 375.00! The "subjective quality" of the language in the translation when it was finished six months later was *perfect*.

Some of us at Logos were recently evaluating our growing list of satisfied customers. A sober, objective, realistic assessment of **why** they are successful, produced some interesting discoveries. There is always a willingness both to invest the time it takes to produce acceptable translation quality and to accept the fact that this will not happen over night. There is always a considerable amount of patience - not just with the MT vendor (although any MT user knows how much of this is necessary) but also with one's own company and colleagues in order to set up the organisation *around* translation in general and MT in particular - and again the willingness to accept the fact that this will not happen over night, either. There is always a nearly religious devotion or dedication to the cause. To summarize: the single most important factor in the successful implementation of machine translation software is not hardware, or software, or networks; it is **you**, the customer. Machine translation **will** work - if **you** want it to.