

Panel

What Can We Learn from Users?

The voices, wishes, and frustrations of experience. This panel is a focal point of the second day of the conference, which is devoted to the practical aspects of MT: the voices, wishes, and frustrations of experience; the hopes and of users; the lessons learned. The panel provides an opportunity to share critical lessons with the audience, many of whom come from a background without users, such as research, or who are users themselves but have little or no understanding of the problems faced in providing the service they want.

Panel members:

Muriel Vasconcellos, AMTA (moderator)

Claude Bédard, TRADUCTIX

Winfield Scott Bennett, Logos

Laurie Gerber, Systran

Edith Westfall, Multilingual Computing Solutions

Learning From Users

Winfield Scott Bennett

Logos Corporation

Introduction

At its heart machine translation is applied computational linguistics. The reality that MT is a practical application is central to all endeavors in our field, wherever they take place. Our purpose is to provide computational solutions for those who use or will use the systems. For the machine translation developers to respond to users' needs and desires quite simply means that we have to let them tell us what to do.

This sounds easy, but turns out to be difficult, primarily because those who develop MT systems and those who use them often have very different views of what MT is about. Both perspectives are valid but may be at odds. The differences often lead to users who believe that MT is just so much technology and developers who view users as Luddites who do not know what is best for them.

While there is neither time nor space to cover the question in depth, I will present two key points which I have learned from users and from which we developers can benefit

Who are "the users"?

For years the machine translation community has labored under the misconception that our users or potential users formed a homogenous group; all we had to do was create software to meet the needs of that group. Reality is that users are not a single group. No one paradigm is sufficient for all the MT users and potential users. In fact one can state that

our users fall into at least four distinct groups (with the understanding that their are the inevitable subcategorizations in each group).

- *Corporate/Government:* Those who use MT as part of publication of documents in a number of languages and/or for information gathering and dissemination. Translation in such a setting may be a barely visible component in the entire process to produce products.
- *Service Providers:* Translation companies whose purpose is to provide translation for those who need them. Generally such bureaus are fairly modest in size. Translation is their product, produced on demand for customers.
- *Independents:* Those translators who work out of their home under short or long-term arrangements with customers. Some have long lasting contracts with corporations; others work piecemeal on whatever jobs are offered.
- *Casual Users:* Those people who need or want translation on an occasional basis.

Of course, there are fuzzy areas in which one group spills over into another, e.g., corporate translation groups who also do service work for outside customers.

The reality we MT developers must face is that our market is much broader and less homogenous than we imagined. For MT to provide solutions to these disparate groups means that we have to listen to each group.

Users' work environments

When I first became involved in MT R&D our model for the users' work environment was basically an in-house group whose task was the translation of technical documentation for publication. Somewhere in our thinking was the idea that the input text was high quality and well-controlled and that such translation was well-integrated and efficient. In addition we saw translation as an event. In 1996 I can state that this model is rarely true, if ever. Our users' work environments do not conform to the MT developers' image.

First, translation *is* a process, not an event. Users may even be fooled into thinking that translation is only a single task, but the truth is that it is actually a process whether done by one person or a group. Translation of parts of a document may actually be done a number of times.

Many of the users have no control over the input text; rather, they are given text—and not all technical documentation—in various formats and with differing quality. This is true for corporate users as well as for service bureaus. Some of the texts may be well-suited for our systems; others are not. The translation process may have to begin with the decision whether to use MT or not.

Contrary to the picture one might have, many corporate users do not have the luxury of an in-house translation staff. Corporate users of Logos, for example, run the gamut from *no* in-house translators to large translation staffs. Outsourcing translation is a reality for many MT users. Management of the flow of text for translation is therefore a significant part of the process.

Fourth, given the realities of release schedules, translation staffs are often confronted with the phenomenon of the 'movable source'. A document is presented for translation at some point, followed at a later point by the first revised version, then a second, and so on. Each revision needs translation of only the new text—but how does one spot it? Similarly technical translation within corporations and for corporations (by service bureaus or independents) often is a matter of translating the documentation for a new model or version

of the product. The technical documentation is mostly the same, but there is a part which needs to be translated. As with the issue of revisions, translating and postediting the entire document is a waste of resources.

Clearly there are other issues; these are a taste of what translation is about.

MT: solution or stumbling block?

If we are to provide MT solutions for users, it is important to learn who they are. The groups noted above each have common and distinct needs. An MT solution for the corporate user may be a partial solution for the independent translators. MT systems must seek to provide solutions for the groups; MT is not a one-size-fits-all solution.

From the users' perspective MT has only one purpose: to assist her or him in the process of translation. The developers' view that MT is an obvious aid may be falsified when confronted with the users' environment. For MT to be of use in the users' environment it has to fit into that environment as seamlessly as possible. The notion that the user must conform to the technology is entirely unrealistic—and unacceptable.

MT systems must conform to the realities of users' work environments and must impose as little change on these environments as possible. This is not to say that MT cannot create some change in the users' environment; rather, that whatever changes come about are natural and not imposed simply due to development decisions.

Of utmost importance is the need for the MT system to fit into the established process in place. Simply offering a technological package without taking the users' process or situation into account is a guarantee that the MT system will be a stumbling block. Developers cannot afford the luxury of dictating a process to the users.

Also important—and sometimes overlooked or ignored—is the need to provide the interfaces and support tools needed for the user to do translation. Ease of use is important, as is learnability. Anything which diverts the user from her or his primary job is an impediment to the translation process. There must be means for the user to incorporate preexisting data, e.g., glossaries or previous translations into the system. If this is not present, the sheer amount of work to enter data can make the use of MT doubtful.

In the final analysis we MTers can learn a great deal from users about what makes our systems truly worthwhile applied computational linguistics. Unless we are willing to listen we are doomed to make interesting technology which has no use in the real world of translation.

Legal Implications of MT On-line

Edith Westfall

Multilingual Computing Solutions

The legal issues surrounding machine translation on-line will need to be defined within the next year. As companies offer MT services on the Web, the opportunities for information to be accessible to more people in more languages will dramatically increase. However, unlike printed information distribution, where publishers and authors have some control

over who translates materials and for what purpose, the Web allows individuals to translate information as they need and want it. More than 25 million people are accessing the Web on a regular basis, with some analysts expecting this number to reach a 150 million in the next three years. Commerce on the Web is expected to increase to more than \$255 billion by 2000. This growth cannot help but increase scrutiny by regulatory agencies and trade groups (such as ASCAP). What will happen when these groups begin to take an active role in the legal questions surrounding MT on-line?

Currently, liabilities can be incurred in several ways. In the case of the Church of Scientology vs. Netcom Online Communications Services, NetCom was sued for contributory infringement [of a copyright] because it did not remove Church of Scientology material which had been illegally posted by a user. This case was settled out of court. In another instance, the On-Line Guitar Archive, an FTP site at the University of Nevada, had to cease operations due to a potential lawsuit by EMI Music Publishing. The site contained lyrics and other information for approximately 15,000 songs. In the latter case, if these instructions had been translated into French and posted, would the on-line MT provider and the web site host have been liable also?

We as business people, academics, and researchers use the Web as a way to gather and disseminate information, but others may use the Web in a less positive way. When looking at the legal issues of MT and the Web, you need to look at not only what is normally done but at what can be done. Will providers of MT be liable for what is done with the translation? Types of liability are copyright infringement, product safety/liability, and possession of banned/discouraged literature.

In our daily work we perform a variety of activities through the Web: looking for information, conducting research, purchasing or ordering items, and posting information. Overall, the goal is to provide and receive as much information as possible. MT on-line furthers these activities. Developers and ISP providers will soon make MT on-line more accessible which in turn raises legal issues. How does work get attributed? For example, a corporate worker translates an article from English to Spanish and posts it on the company's intranet—who owns this work? Are royalties owed? If the author or publisher has not given the company permission to post the article does the company have a right to MT it? What counts as translation for distribution? Who owns the rights to MT output of on-line text? Does the author retain rights over the text? Is anyone liable for an incorrect translation that leads to a safety problem?

One answer is to have an MT Ready or Translation Ready bug (stamp) similar to the union bug. Adding this stamp to a site implicitly gives the browser permission to use MT. On the other hand this is a daunting task because there are millions of web pages.

The Internet has been promoted as a vehicle for communication but the Web has evolved beyond that to encompass commercial transactions. As a communication device the US Postal Service is involved; as an international commerce facilitator, the US Customs Service is concerned; and as a domestic commerce facilitator, the FTC as well as other agencies potentially has a role. These organizations have broad powers over communications and commerce and have not yet written the regulations and standards for the Web. Should they? What is their role to protect and regulate?