MACHINE TRANSLATION IN SOUTH-EAST ASIAN MINORITY LANGUAGES

Stephen Beale
Thammasat University / Summer Institute of Linguistics
Language Research and Development Project
Thammasat University Bangkok, Thailand

1.0 Minority Language Translation Needs.

Why minority languages?? Most people recognize the validity of a machine translation (MT) system between two large language groups which share a large volume of textual resources (i.e. technical manuals, information services, scholastic research, etc), but how can one justify the large expenditure of resources necessary to establish a translation system that is aimed at a relatively small population with relatively small translation needs? Is a system like this practical?

No. In one sense it is not. It is certainly not practical in a business sense, at least not in the short term outlook. To implement a comprehensive system for an entire country, typically linking together four or five main language groups along with all the related dialects, would require many years of linguistic research. And the benefits are not completely obvious.

But the minority language issue is a complex one in any country. There are many reasons why an automatic translation system targeted at minority languages could be very useful - both for the national government and the local people. This paper will seek to outline some of the positive results a policy of translation for the minority languages might have, and to present two MT systems that might be used together to implement such a policy.

1.1 An Alternative to Integration.

An effective translation program between the national language and the various minority languages could potentially offer many advantages. In general terms, it could produce many of the benefits of linguistic "integration" or "absorption" without creating the unwanted side-effects involved. What
exactly is meant by "integration," and what are the benefits and side-effects associated with it?

Integration, in the context of this paper, would produce a situation in the country where every citizen used the national language with equal facility and expertise as any other citizen (discounting the normal variations in intelligence and schooling found in "normal" communities). Every citizen would have equal access; no citizen would be at a disadvantage because of the choice of national language.

Integration is occurring in many countries around the world. In some places it is an official policy vigorously enforced. In other areas it is a naturally occurring process brought on by socio-linguistic factors. Wherever it occurs, the ultimate result will be that the entire population of the country will speak the national language fluently (and often exclusively).

The benefits of such a situation are obvious. Communication and written resources could flow freely to all parts of the country. No group would be isolated from any other groups. The mistrust that is often present between the national culture and the minority groups would be greatly lessened as communication is facilitated and misunderstandings reduced. Because of these and other results, the national government would not have to allocate special resources to the control and aid of the minority groups.

The side-effects of such an integration are, however, substantial. From a cultural and anthropological point of view, the suppression of the minority languages along with their distinctive cultures would be a great loss to the country. It would be naive to believe that integration could be achieved without the eventual collapse of the minority cultures. Perhaps a more pressing problem would be the unrest and revolt a policy of forced integration would bring. Minority groups (indeed any group, large or small) will not willingly accept the destruction of their language and culture. Many of the world's greatest conflicts (past and present) concern the relationships between national cultures and minority groups (note the recent problems in Yugoslavia and the Soviet Republics).

Of course, a system of gradual integration could be implemented, but such a system would need many decades to take hold. In the meantime, all the problems of a multi-lingual society must be dealt with. And whether gradual or rapid, the process of integrating a people group into the national language is, practically speaking, a very hard goal to achieve.
1.2 Some Benefits of Translation.

A multi-lingual translation system could, then, offer many of the benefits of integration without producing the drawbacks. Politically, communication between the government and the minority groups would be dramatically improved. As for the minority groups themselves, they would have much greater access to the resources of the national language. As these resources become available, many of the problems faced by these groups could be overcome. As a result, the status of the minority groups (in their own eyes as well as those of the country) would increase, making them more productive and important members of the country.

As noble and exciting as these results might seem, they might not even be the most important. The psychological effects on the population might eclipse them. A full scale development program of this sort by a national government would send the following message to groups all over the country: "we care about you, and we are willing to work at making you an important part of our country."

1.3 Some Useful Translatable Materials.

What types of translated materials would be useful for minority groups? The four areas of health, education, employment and politics summarize the important possibilities. Health problems often plague minority groups. Literature promoting and teaching good health practices would be invaluable. Information ranging from waste disposal to disease treatment could eliminate many of a group's problems. A book such as Where There Is No Doctor [Werner 1977] (various Asian editions are available) could be used.

Educational materials could be translated into the minority languages. Even if eventual integration was a goal, a translation system could help the process by providing a solid education at the lower levels. It is generally accepted that education in a national language will proceed much better when supported by a base of "mother-tongue" education. School textbooks and other appropriate reading materials for children could greatly enhance the education of minority groups.

Employment helps could also be a potentially fruitful source of translation materials. Included here would be information about farming methods, as well as literature related to other professions common in minority groups. Vocational training could also be of benefit as these groups move into greater contact with the national society.
Political information would help keep the minority group in touch with the rest of the country. Election materials and laws, along with a host of other materials in their own language, would greatly benefit minority groups. Translating a daily newspaper might be one of the most significant contributions. A newspaper is a source of all sorts of information, including all four of the areas described above.

Many other kinds of literature, from religious to pleasure reading, could be translated too. And, as technology grows, more sophisticated systems could be implemented. In 50 years (or 100?), systems that input and output speech may be developed. Accessing information from all over the world may become as common and as easy as going to the library.

2.0 Machine Translation Methods for Minority Languages.

Two approaches to machine translation are viable for reaching minority populations. To service languages that are closely related dialects of the national language (or are closely related to some other minority language for which translation has already been done) a dialect adaptation system (i.e. CADA) will probably be sufficient. This type of system transfers the surface structure features of the source language (including morphology and syntax) to the corresponding surface forms of the target language. Basically, this system will make all the needed changes to the vocabulary, and will rearrange syntactic and morphological structures as necessary.

For more distantly related languages (for example Thai and a Mon-Khmer or Karen language), those types of operations on the surface level alone will not be sufficient. Semantic analysis of the source text will need to be done, with the resulting semantic representations being transferred to the target languages. This approach is to be implemented in the TU-SIL Translation System.

2.1 Computer-Aided Dialect Adaptation (CADA).

The CADA approach, as its name implies, is useful for translating between related dialects. This raises an important question: how close do two languages have to be to be related'? Would Thai and Northern Khmer be considered "related"? This question will need to wait for further experience. Finding out how far one can push CADA and still obtain acceptable results will be a priority of the upcoming research.