Jens Allwood
Department of Linguistics
Göteborg University

LANGUAGE TECHNOLOGY AS AN AID IN PRESERVING LINGUISTIC DIVERSITY

There are between 4 000 and 8 000 languages in the world. The reason a more exact figure cannot be given is that linguistic factors are not sufficient to define a language. Rather political factors are intrinsic in the concept of language. Today a dialect (language), tomorrow a language (dialect). As an example, consider the situation in Yugoslavia. 20 years ago there was an attempt to make Serbo-Croatian the national language. Today linguists are helping to make Serbian, Croatian and Bosnyak into distinct "languages".

Leaving difficulties of definition aside, it might not be unreasonable to claim that there are say 6 000 languages in the world. Many of these are disappearing or are threatened by extinction. The reason for this is that they often are spoken by a small, aging population. Some languages today are used only by 1 or 2 speakers who are all over 60 years of age.

If we want to preserve the linguistic diversity of our planet, the key factor is the usability of the languages. There have to be opportunities and desires as well as needs to use the languages. The desires and needs are created by, for example, loyalty, tradition, group membership and by the possibility of sharing thoughts with others. However, if this possibility is really going to count, it has to come into play in satisfying practical needs (such as word, food, career etc.).

The opportunities for using a language are created among other things by the technological support for communication in the language. This is where language technology and dialog systems come in.

Language technology can play an important role both to preserve and to support active use of the world's languages. It might be argued that even if we cannot retain active use of all languages, we should at least try to preserve a record of them for future generations. Today we definitely have the technological resources for this.
The   basic need here would be to create corpora of written (even today there are languages that don't have writing systems) and spoken language. Preferably
the corpora should be in multimedial form so that bodily communication and typical situational contexts would also be preserved. Since so much of language technology depends on written language, multimedial corpora should also, where possible, be accompanied by transcribed versions. Once fairly large scale corpora have been created, they can then later be subjected to many different kinds of linguistic analysis.

If we have a more ambitious goal than preservation, there are even more things that could be done using language technology. But it should be said that large corpora of spoken and written language remain a basic resource out of which other tools can be developed.

Let me end by giving a list of some examples of the support that can be given.

1) Linguistic interfaces for operating system and communication program
2) Word processing systems (with dictionaries, support for spelling, hyphenation and grammar)
3) Speech analysis and speech synthesis
4) Dialog systems
5) Information retrieval via translation into larger languages
6) Multimodal communication systems

The list can be made much longer but points to real and important tasks for all of us that are involved in language technology and dialog research.It is to be hoped that more linguists interested in dialog systems and language technology become aware of the field of language preservation as an interesting area of application for language technology.