

MULTIMEDIA TECHNOLOGIES FOR THE USE OF INTERPRETERS AND TRANSLATORS

By
Angela Carabelli
SSLMIT, Trieste

The availability of teaching materials for training interpreters and translators has always been an issue of unquestionable importance. Too often, however, experiments are carried out in an exclusive or unique environment with ad hoc resources, which are rarely made available for widespread use. This may lead to two different consequences: on the one hand, experimental results could prove to be statistically invalidated because testing is often carried out using little material with a limited number of subjects; on the other hand, once the tested approaches have proved their validity, due to the lack of large-scale distribution of the resources used, students cannot always benefit from their results.

The complex nature of interpreting and translating skills has many implications, not only in terms of assessing valid teaching and training methodologies but also in terms of making the proper instruments available to students.

In most teaching environments, students often have difficulties in finding material suited to their particular needs due to the lack of facilities for the targeted search of specific texts. In addition, any possible audio recording for student use is hardly ever accompanied by corresponding written versions and is made available in a library-like fashion (with all its implicit limitations).

The question is how can these resources be organized and made available in a way which would guarantee greater efficiency, flexibility and autonomy as an aid to students' training? In conceiving an information system capable of catering to these specific needs, the idea of resorting to computer technology was deemed the most feasible solution. If, in the past, the management of multimedia archives created certain difficulties, the rapid advancement of technology and the recent boom in the use of Internet resources have paved the way for a new era of highly efficient services.

Unfortunately, many software tools available today are considered inadequate because they are often developed without the constant feedback between programmers and end-users – a vital condition to determine the success of any experimental project. During the various phases of implementation of an optimized data structure for the management of resources, the author aimed to avoid this pitfall. For this reason, the description of implemented technologies is better given from a "user's" point of view, i.e. following the levels of manipulation which can be reached easily by any "computer-shy" user, and not

including the specific implementation details, such as programming languages, as they are strictly pertinent to the field of software engineering.

The project of realizing a data base of resources for interpreters and translators starts with the analysis of the software applications available today, both in the commercial as well as in the scientific and research sectors. A detailed study of the virtues and vices of such products has highlighted two fundamental problems:

- Interpreter products have yet to be implemented or are not currently available: the majority of products have been developed with the objective of providing an aid to language learning, often at the basic skill level
- All of the products that are currently available have been conceived for single use only, on single work stations and without the possibility of interaction among users.

The need for a large number of resources pertaining to different specialized fields, and their availability for an indefinite number of users, are thus factors of great importance. In addition, it must be observed that today's available resources are limited to the audio source and do not include extra information that could prove strategic not only for students, but also for the preparation and study of professionals and/or teachers: such information may include indications regarding context, teachers' suggestions for particular collocations, explanations of acronyms, the written version of the text at hand.

Giving access to resources to a large number of users allows for strong interaction, the exchange of information and, consequently, the realization of a "live" and self-fed information system.

In other words, the analysis and research carried out by students themselves during practice and their simultaneous use of resources can lead to the retrieval of new information and to the building of general glossaries and references.

The first step to take in order to assess the feasibility of creating a data base is the analysis of currently available technologies and their possible exploitation for the objectives set out to be achieved.

A discriminating factor is the fast spreading use, over the past recent years, of multimedia technologies for resource management. From this point of view – two features are highly indicative:

- 1) the strong homogeneity of standards for audio and video resource archiving
- 2) the rapid fall in costs for the necessary hardware support of multimedia applications.

The use of sound cards for digitizing and reproducing sound, in particular, allows the reader and/or interpreter to record his voice and subsequently archive the newly created file on the appropriate support. A number of tests carried out with low cost commercial cards has shown that the processes of reproduction

and recording of digital sound may be activated simultaneously, thus enabling the individual to practice simultaneous translation. The availability of several in-going and out-going lines, in addition, has made the configuration of a complete workstation based only on a Personal Computer feasible. What is needed is a microphone and ear phones respectively for recording the interpreter's voice and hearing the original text, as well as a traditional analogical device (for instance, a plain tape recorder with magnetic audio cassettes) to digitize already existing tapes or to record the sound of a digital resource onto an audio cassette.

As far as the more traditional form of archiving is concerned, that is, the archiving of written texts, the evolution of programming languages, the creation of new standard formats and the progressive improvement of computer performance have enabled common users to rely on the highest of quality levels. Texts may also be stored in different formats, all compatible with each other, including particular layouts and, as in the case of HTML language, references and connections (hyperlinks) to other resources. Considering the level of current technologies in the field of scanner devices and programs for optical character recognition (OCR), the written version of the original texts (where available) may be archived together with the audio and/or video versions.

Distributing information to a vast number of users is no longer a serious problem as the growing use of Internet-based technologies has effectively shown the validity and the power of telecommunication networks. In particular, with the objective of enabling a large number of users to gain access to digital archived-resources mainly within the same organization, but also potentially in different organizations, an information system that is active both at a local (Local Area Network) and a wider level (Metropolitan Area Network or Wide Area Network) must be outlined. The need to make resources uniformly available and to maintain a strong level of access control (the main purpose of the data base is didactic, and the resources made available must therefore be subject to supervision by the teaching faculty) require the centralization of a service at the level of a central Server, to which users may connect.

The setting up of an extremely modern network infrastructure at the School for Interpreters of Trieste (realized with structured cable) and its connection through the Metropolitan network of Trieste to the general University and GARR (the MURST national scientific network), suggest the feasibility of an open information system based on the TCP/IP protocol, allowing local and remote communication.

As regards the technology used for realizing software applications for both a central Server and remote work stations, two fundamental strategies may be identified. They are based respectively on the use of what are known as Web technologies and on the development of dedicated software.

In the first case, the Server must activate a Web service with which users may communicate by using a simple Web browser, while dedicated software requires the realization and distribution of applications developed exclusively for access to resources.

Both approaches have their advantages and disadvantages and having to choose either one or the other would prove to be very limiting. The advantages in using a Web Server are clear as they derive from the use of public or freely distributed software, available on any Software/Hardware platform, and there is no need for the installation of particular products on the workstations. Clearly, with this solution the spreading of access could be considered global, in the sense that any Internet user would be able (if, of course, access is authorized) to connect to the Server and use the resources available. In the case of a dedicated application, access to the system would require the implementation of specific software on each work station, and would therefore not be available on such a wide scale. On the other hand, the functions of the user interface realized with Web technologies could never equal those of dedicated software, especially when dealing with the management of multimedia resources.

When considering the pros and cons of the two possible solutions, the decision was made to create a base system that was extremely flexible and that could be easily interfaced in both ways, thus offering complete and optimized access through a dedicated application in the Windows environment (certainly the most widely used in schools) and access to its main functions, while distributed on a larger scale, through a Web service.

The management system for a data base is the cornerstone of the project's general philosophy, because it constitutes the working center of all of the transactions and operations carried out by the users. The need to manage a large amount of data (the objective is to provide interpreters with a large selection of resources in all sectors and in all of the different source languages), which is well-structured (to create a system that allows for the expansion of resource-associated information and its easy retrieval) without a doubt requires the adoption of a data base management system – DBMS, rather than the structured archiving of plain resource files on disks.

All of these factors must be supported by a management system and by an especially designed data structure, foreseeing the need for:

- reorganizing managed information
- integrating/modifying the implemented attributes
- reclassifying managed information
- modifying/introducing control rules

As may be observed, the system must be realized with technology that offers considerable flexibility and allows for any necessary radical restructuring, minimizing as much as possible the need to apply such changes to all of its

constituent parts. In other words, all of the functions subject to reorganization must be managed at a centralized level by the DBMS and must be deferred as little as possible to access systems adopted by users. The best way to satisfy these needs is to implement a Relational Data Base Management System, (RDBMS) in a Client-Server architecture in which the realization of integrity functions (verification of the correctness of information, the verification of accesses, referential integrity, targeted research procedures, etc.) is carried out directly on the Server, limiting the role of the Client to plain instruments of system interfacing.

Defining the data structure entails identifying fundamental entities for information management. To this end, the main object of the information system, that is, the texts in their original language, were taken as the starting point, while gradually, according to needs, a series of new accessory or auxiliary entities were then defined.

The main archives of the system must allow for resource management by interpreters: such resources are basically characterized by a series of attributes that indicate their origin and language, by a few secondary parameters (such as context, background information, etc.) and by fields for the archiving of data related to digital audio and video and of the written version of the text.

The need to give rapid and limited access suggested the realization of a series of archives based on the classification of different texts. Their organization is characterized by entities that follow the structure of the environment in which they are formed. Archived resources may be associated, for example, to courses of a specific faculty for which they are considered valid. In this way it is possible to indicate whether a resource is pertinent or advisable for the interpreter's/translator's practice within a specific course, allowing subsequent users to pick out from the data base the exact resources that suit their needs.

In the latter case, the solution would be to create a hierarchical table of categories that enables us to define actual classes for the organization of the texts: during referencing, one may access all of the resources belonging to a certain category, automatically including in the selection all of its subcategories as well.

In order to allow for an intense and facilitated use of the system, as well as the exchange of information and translations, an archive for the management of translated texts has been created. This solution is part of a larger project – IRIS (Interpreters' Resource Information System) – which comprises and exploits the above-described multimedia technologies, and was conceived mainly for academic environments. With such a system, archiving translations that can be rendered in both digital sound (simultaneous, consecutive and sight translations) and written form, must be done according to the type of user connecting to the

system or to the preferences indicated upon accessing the system. Specifically, the translations were necessarily divided into three standard classes:

Private Translations

translations that are stored by the users, that remain at their disposal, possibly for a limited period of time, and that are available only to the author of the recording, or possibly to those users who have been authorized by the former through a message

Shared Translations

translations stored by users that are made available to *other* users

Published translations

translations that, for specific didactic use or quality, are made available to system administrators or by members of the teaching staff and are accessible to all users

It must be observed that this type of archiving makes shared resources available for comparisons and checks among different students. It also makes student translations visible to teachers for evaluation, while at the same time maintaining, full control of the management of the data base's resources.

In order to facilitate the interpreter's and/or translator's work, a system that allows the user to attach notes to the original text has been foreseen. These notes may include essential indications as to the meaning of acronyms or particular initials found in the text, or special suggestions for specific collocations.

To this end, the term Pointer is used in association with a word, an acronym, or more simply a piece of the original text (even an entire sentence, if necessary): each Pointer is archived and codified in a general archive system, together with its written explanation. If one or two Pointers coincide but have different meanings, users may archive all of the pertinent explanations by attaching a code number to each Pointer.

Two main advantages can be drawn from this approach:

- The ordinary use of Pointers allows for a recycling of definitions and explanations already used in a number of different texts, facilitating data management and reducing the amount of space used in the databank's memory.
- The associations between Pointers and texts may be referenced in the reverse manner, allowing users to search for all of the texts in which a given Pointer is used.

In the specific case of original texts available in written form, the retrieval of Pointers was made easier through links to the parts of the text in which they are used: these links, highlighted in the original text, have been called HyperHints.

The use of a system by a vast number of users requires a specific management strategy for authorizations that make it possible to differentiate the possible types of operations. For example, the realization of new resources can be limited to only senior students or the archiving of translations can be reserved only for teaching staff.

A number of optional features have been designed for future enhancement of the system. It has been conceived in a way so as to create not only a teaching instrument capable of meeting the needs of trainees during the learning processes of translation, but also a more efficient means (supported by the necessary research of authorized personnel) for monitoring the various phases of users' progress and for assessing the quality of their work.