

## Language as Cultural Heritage – promoting research and public awareness on the internet

Stephen C. Levinson & Peter Wittenburg  
Max Planck Institute for Psycholinguistics  
Nijmegen  
The Netherlands

### **Abstract:**

*The ECHO proposal aims to bring to life the cultural heritage of Europe, through internet technology that encourages collaboration across the Humanities disciplines which interpret it – at the same time making all this scholarship accessible to the citizens of Europe. An essential part of the cultural heritage of Europe is the diverse set of languages used on the continent, in their historical, literary and spoken forms. Amongst these are the ‘hidden languages’ used by minorities but of wide interest to the general public. We take the 18 Sign Languages of the EEC – the natural languages of the deaf - as an example. Little comparative information about these is available, despite their special scientific importance, the widespread public interest and the policy implications. We propose a research project on these languages based on placing fully annotated digitized moving images of each of these languages on the internet. This requires significant development of multi-media technology which would allow distributed annotation of a central corpus, together with the development of special search techniques. The technology would have widespread application to all cultural performances recorded as sound plus moving images. Such a project captures in microcosm the essence of the ECHO proposal: cultural heritage is nothing without the humanities research which contextualizes and gives it comparative assessment; by marrying information technology to humanities research, we can bring these materials to a wider public while simultaneously boosting Europe as a research area.*

### **1.0 The hidden languages of Europe – the example of Sign Languages**

ECHO has the ambition to simultaneously advance access to the cultural heritage of Europe and to promote and network Humanities research across the continent, through development of the technology to make Humanities research on cultural heritage available on the internet. One of the richest elements of the cultural heritage of Europe is, obviously enough, the major languages of the member states, and their accumulated national literatures. This is well enough appreciated that on-line access to such resources may seem hardly important. Nevertheless, which of us can read Beowulf or the Norse Sagas or Petrarch in the original? Here the provision of images of the early manuscripts, line by line glosses, on-line dictionaries, and cultural footnotes could in fact provide a wonderful resource, allowing wide public access to the historical roots of all the languages we share here in Europe, with obvious educational applications. The technology required here would essentially be the same as that required for displaying and interpreting materials from the History of Science (see the relevant section of this booklet).

There is another side to the linguistic heritage of Europe, a hidden side. There are disappearing dialects, and the languages of ancient minorities, from the 13 Romany dialects to Breton to Fries, on which information could be provided. Many of these ethnic minorities are now dispersed around Europe and beyond, and the possibility of reconstructing a “virtual community” – to which people could contribute spoken language materials – would be very attractive to the public. One could for example have a Celtic site, inside which one would find Irish, and therein one could find recordings and films of spoken Irish from different generations, collected not only from Ireland but from all the destinations of the Irish diaspora.

There are also the many languages of immigrant communities which now play such an important role in our economic and political life, and indeed in our demographic future. The

provision of materials on Turkish, the Caribbean Creoles, Tamil, Chinese, Vietnamese and all the other languages of our immigrant communities would be much appreciated, and would help to network the wealth of scholarly knowledge about such languages in the Universities of Europe.

But let us give the most striking example of a hidden linguistic richness right around the continent, about which the proverbial man on the street knows next to nothing. In Europe there are many languages that are all around you but which you have never heard. They are the Sign Languages, languages used by the profoundly deaf and their relatives. In every deaf community, Sign Languages are the natural, preferred means of communication. Much of the public imagines that such Sign Languages are just gestures, some primitive universal iconic medium, not realizing that each of the Sign Languages of Europe have evolved over centuries into real languages with the full expressive power of any of the major spoken languages. They have their own distinctive words, and their own syntax – their own ways of combining words into sentences. Each of them is the focus of a rich cultural system, and they have unexpected cultural connections across and beyond the member states. Thus Austrian Sign is related to Russian sign, French Sign to American sign, but British Sign is unrelated to American Sign, and so forth. Some member states have more than one Sign Language, as in Catalan vs. Spanish Sign, or Belgian French vs. Flemish Sign (the latter not closely related to Dutch Sign incidentally). Altogether, there are at least 25 Sign languages of Europe, and at least 18 indigenous Sign Languages within the existing EEC.<sup>1</sup>

There is an important ethical dimension to the appreciation of Sign languages. The history of Sign languages in Europe has been a roller-coaster ride, from high points in the Enlightenment to the lowest depths in Nazi-occupied Europe. Much of the history is tragic, with repeated acts of suppression of Sign in favour of the dominant spoken languages taught through finger-spelling and attempts at vocalization. New justifications for the suppression of Sign have been based on medical advances (cochlear implantations), but Sign remains the favoured expressive medium for all deaf people. Member states still take different attitudes to their Sign languages, and only three (Greece, Finland, Portugal) give them constitutional recognition despite two resolutions of the European Parliament (1988, 1998) encouraging this. There are thus important public policy implications here.

There is tremendous public interest in Sign Languages, which is currently unsatisfied. There are frequent enquiries from the parents of deaf children, or from those who have a deaf colleague or in-law, betraying both a curiosity and concern and a deep public ignorance about the essential facts.

Sign languages present a major technological challenge. Needless to say, remote communication by Sign is currently problematic. In Texas there is a pilot experiment with video-phones, and spoken-phone to video-phone links provided by bilingual operators – this has proved to be a great success. In Britain there is a project to create a ‘virtual signer’ – a computer driven animation which will take English as input and produce British Sign Language as output, in order to facilitate communication between the hearing and the deaf. Research on Sign Language is equally technologically demanding. We must have ways to annotate visual recordings (e.g. with translations into written language), and search and retrieve segments of such visual episodes. This annotation should be carried out collaboratively between partners

---

<sup>1</sup> For current information on the status of the Sign Languages of Europe see Newsletter March 2001 of the European Union of the Deaf, <http://www.eudnet.org/shownews>. There is no accurate demographic information on Sign Language users, but a conservative estimate would be that there are half a million in the EEC. The enlargement of the community will substantially add to these figures. The authoritative catalogue of the languages of the world, the Ethnologue, lists 122 Sign languages worldwide, and comments on most of the European ones “little researched”.

working at remote sites and other researchers should be able to add comments in a flexible way. Preliminary solutions are sketched below.

Sign languages are scientifically important to our understanding of the very foundations of human language. No other animal can displace its communication system from one modality to another in the way that human language can be displaced from the vocal-auditory channel to the manual-visual one. Language has a deep biological and psychological basis – but the very existence of Sign Languages challenges our current understanding of these foundations. For example, language is largely in the left-hemisphere of the brain, but spatial cognition is largely processed in the right hemisphere – so what happens when a spatial language is processed by the brain? It is thus scientifically imperative to understand the nature of Sign Languages at a greater depth than we currently do. The first thing to do is to deepen our understanding of each of these languages. There are some strong research traditions, for example in the national institutes of the deaf of the various member states – but there are also many member states where no such scholarly tradition has yet been established. But the *comparative* study of the Sign Languages of Europe has hardly begun. So here is a natural network waiting to be established across the continent – a network which would bind together scholars who currently work largely within national boundaries, and without the benefit of the technology that we could bring to them. For the study of Sign is technologically demanding – it requires from its very nature a multi-media approach, a method of handling video documents with rich annotations and full retrievability. Most current work uses ad hoc means and relatively primitive home-grown technology.

So here would be an immensely valuable project – to introduce to the citizens of Europe these unheard languages in their midst, to show them how conventionalized and expressive they are, and how they have the status of real languages, to explain historical relatedness and cultural differences. We would do this first in the following way. Each of the 18 languages of current member states – from Portuguese Sign, to Greek Sign, to Belgian Sign, to British Sign – would be represented in the following way:

- Historical pages, detailing e.g. the eighteenth century pioneers in the various countries
- A video clip retelling the same story from a wordless picture book in each languages, allowing visual comparisons between each and every language
- A large set of words in each language – e.g. the word for fire, man, tomorrow – showing how different the different Sign systems are, and relatedness where it occurs
- Conversational exchanges in each language, to show that these are living languages used in ordinary face-to-face communication
- Folklore and poetry in Sign, to illustrate that there accumulated cultural riches – a ‘manual literature’ paralleling the ‘oral literature’ of traditional spoken tongues
- More detailed linguistic material, arranged on a comparative basis, for each of the Sign Languages.
- Using the tools described below, internet users could find the video sequences of interest to them, search for specific features and compare e.g. the same sentence in three different Sign languages on the same screen.

This kind of information could be assembled from across Europe from the national researchers already at work on the local Sign systems.

Such a program would open up a hidden part of European heritage to the wider public. It would help us all to appreciate the achievements, the culture, of our silent fellow citizens. But it would have real scientific value too – the fact is that no such comparison has ever been done, and we have no clear idea about what the results would look like. The results would be of great interest to students of language and psychology. It would be a first – yet another example of Europe taking the lead in an area rich with cultural appreciation, policy implication and ethical dimensions.

## **2.0 Tools for multi-media handling of spoken language, Sign and other cultural performances**

Cultural objects – scanned or digitized – are of little intrinsic interest without the information and scholarship that sets them in context, explains their importance and compares them to others. All this information is linguistic – and linguistic tools are therefore fundamental to the entire ambitions of the ECHO project: they play a crucial role in retrieval of information, making it accessible across the national languages, and so on. In addition, language can also be the object of attention, as when the object is an important historical text. Here, we focus on just one kind of language material, recordings of linguistic performance unfolding over time. The tools required here are specifically related to the dynamics of events, and we will describe these below.

First, however, imagine users of the internet wishing to find out something about the languages of Europe. They can click on a map of Europe and then successively narrow down the area they are interested in, and the language they want to hear and see – finally a list of different kinds of recorded speech events are offered to them, and through successive choices they arrive at a specific film, of say a person describing the local city in Dutch Sign Language or the local dialect of Romany. They can get the simultaneous translation in different major languages, or they can search for a specific feature, like a greeting or the word “Hello”. (Maps of course can be more than a way to find the data – they can also offer, for example, historical overviews of the extent of different language areas over time.) Alternatively, instead of using the map, they can browse in conceptual spaces and/or search with keywords ‘description in Dutch Sign’ or the like, and rapidly they find the relevant excerpts.

This multidimensional mode of access requires underlying structure of course. We have developed a “Browsable Corpus” system which delivers all this, while at the same time facilitating the task for the provider of the information through a specific Browsable Corpus Editor which structures the input information and guarantees adherence to the constraints of the keywords required for efficient retrieval (the set of ‘metadata’ attributes). The metadata set used was developed within a European initiative and supports description elements allowing researchers and also the interested public to enter detailed queries. For the general user metadata harvesting with the help of OAI based protocols and Dublin Core elements is supported. The Browsable Corpus not only navigates the user to the objects of interest (assembled as a single coherent set of documents and data, though they may in fact be distributed around the internet), but also automatically launches the tools appropriate to the kind of data. This kind of structure in fact is essential to the entire ECHO project, and we plan to make the Browsable Corpus cover all the major types of data covered by research in the humanities, whether pictures, music, ancient text, living language or for that matter any kind of performance.

So, the Internet user navigates to the linguistic performance of interest. There we launch a viewer to see the digitized film of the event, by streaming of media fragments and synchronized annotations. We have developed a specific tool set, EUDICO, which is designed to input, view and analyze annotated multi-media records. Here we have had to do the underlying conceptual work: any event unfolds over time, and annotations need to be locked to a time frame or to already existing annotations. But some annotations are dependent annotations, and inherit the time frame of the higher order annotation, while others are overlapping in time. Again others may refer to annotations on the same or other tiers. All this requires an underlying computational structure which can also be exported to open XML-based formats. EUDICO is built on such a structure, and consequently you can search video-records for specific attributes (say the Dutch Sign for ‘girl’), and immediately see just the relevant sequence, and all the time-locked annotations (translation, linguistic annotations, etc.). EUDICO has general potential application to a video-record of any cultural event or performance of any kind – for example,

the user could find and compare the styles of football playing across the member states, or study techniques of traditional boat building (if they have been documented by video recordings). But in fact EUDICO has been developed so far specifically to handle linguistic and gestural material, and this provides some insight into the way in which a specific research goal in the Humanities can drive technological development. Language-special features include the display and analysis of the speech signal, the specific annotation tools, and the kind of time-bound manipulations available from the interface. For example, for the study of gesture and Sign, one needs to be able to lock into two or more very small time windows enumerated in milliseconds, so that overlapping events can be characterized both independently and in relation to one another (say speech and gesture, or manual Signs and facial Signs).

The kind of tools sketched here need development in various directions. The Browsable Corpus needs to be developed to handle different kinds of object material, to support flexible user-definable hierarchies and to have general procedures for finding and launching analysis tools in a distributed environment. The EUDICO event analysis tool needs to be generalized so that annotations can be made onto the same general database over the net, so making it possible for remote research partners to collaborate on the annotation of specific events and to add comments to a particular time-locked series. These are quite demanding technological developments.

## **Conclusion**

This short contribution sketches an approach to language as cultural heritage, within the framework of ECHO. We have suggested that the many distinct Sign Languages of Europe are part of the hidden cultural heritage of our continent, which can be unlocked by provision of the technology that would make the systematic comparison and collaborative research on these languages possible – while at the same making them available to the general public on the internet. The required technological tools are quite clear and circumscribed, but would have wide application. The Sign Language project can stand as a microcosm for the ECHO initiative: let us use the technology to simultaneously satisfy public interest in our cultural heritage and boost research and collaboration in the Humanities – in doing this we will drive the technology in demanding and profitable directions, and help to construct a European research area in the Humanities.