

# **Implementation and evaluation of an advanced tool for the annotation of authentic audio resources in second language instruction.**

D. Donald <sup>1</sup>, R. Gianninoni <sup>2</sup>, J. Goldman <sup>3</sup>, M. Lister <sup>2</sup>.

<sup>1</sup> *Glasgow Caledonian University, Glasgow, Scotland*

<sup>2</sup> *CILTA, University of Bologna, Italy*

<sup>3</sup> *Northwestern University, Illinois, USA*

## **Abstract**

Project Pad, an open-source endeavor under development at Northwestern University, is an integrated web-based set of tools for searching, organizing, annotating and analyzing objects from digital repositories, for integrating those objects and analyses into web-based projects and for group exchange and decision making. It supports synchronous interaction between users and interaction with media objects themselves, such as the segmentation and annotation of streamed audio or video. These annotations can be permanently stored and searched transversally. Techniques for the integration of Project Pad into legal English courses at CILTA, the University of Bologna language centre, are presently being tested. Analysis of student feedback in the form of reports and recordings, as well as the XML data logs, will be used to establish criteria for the evaluation of students' linguistic choices during annotation, their behaviour and expectations when dealing with unfamiliar audio content and for the development and assessment of micro- and macro-listening comprehension skills. It is also hoped to gain deeper insight into modes of interaction between language learners in relation to streamed media content.

*Keywords: authentic multimedia resources, annotation, digital, collaborative commentary, digital collaboratory, interactive, open source, Project Pad, Spoken Word, streaming, repository.*

## 1 Introduction

Digital learning environments are educationally effective if they create engagement at both the perceptive and the communicative level. The manipulation of authentic multimedia resources using advanced collaborative tools would be one means of reinforcing this process and making an effective contribution to the transformation of traditional teaching methodology, as well as encouraging the users of e-learning systems. Language educators should strive to create conditions in which language learners may search and access a multitude of sources, to store, link and organize information and to engage in critical comment with their peers about the chosen media. Access to significant public audio collections as described in the executive summary spoken-word audio [1] is crucial to this enterprise. The collaborative partnership between Northwestern University, Glasgow Caledonian University and the University of Bologna intend to bring to maximum fruition the experience of the Spoken Word project [2] in applying digital library science to education and experience of CILTA in applying ICT to language learning.

## 2 A three-tier system for collaborative learning with multimedia resources

Our aim is to establish a digital language resource and ICT infrastructure capable of supporting innovative teaching models and practices, firstly by facilitating access to large-scale digital audio collections, such as the OYEZ Supreme Court database [3] or BBC archives, under the auspices of the Spoken Word project, and secondly by enabling learners to interact with these resources. It is furthermore hoped to demonstrate the effectiveness of such a structure in terms of enhanced learner autonomy and learning effectiveness.

Over the previous six-month period, we have laid the foundation for three distinct technological levels needed to create an integrated environment for the individual or collective manipulation of audiovisual resources and for their organisation and online distribution. The physical level, that of storage and streaming delivery is described in detail by Lister *et al* [4]. The second, logical level, concerns media archiving, description and access. The choice of *Fedora* [5], a system which not only allows for the description of multimedia objects but also the definition of associated Web services with *media (Behavior Mechanism)* and method of presentation (*Extensible Disseminator*), is outlined in detail by Gianninoni *et al* [6]. The third, application level, is where we find the most important advanced web applications. In 2004, two ALTAIR Legal English modules at the University of Bologna were chosen to test the use and utility in language instruction of one such open-source application.

### 2.1 Interaction with media and sharing of resources by users: Project Pad

Project Pad is a web-based system for scholar, teacher and student collaboration and media annotation. The system makes it easy to enrich content, learning and

study environments. The system integrates tools: for searching for - and organizing any digital object from - digital repositories; for annotating and analyzing the objects; for integrating the objects and analyses into a scholar's (or a student's) web-based projects; and for group exchange and decision making. How this might work in practice is described in detail by MacWhinney *et al* [7]. Project Pad enables a new paradigm for rhetorical and discourse analysis by providing researchers with ongoing, real-time access to a core set of data that form the empirical basis for their work by providing tools that allow immediate communication and exchange of analysis. Project Pad makes it easy for a work group to collaborate and lets the group members develop and respond to each other's insights and observations.

Collaborative systems have made great contributions in areas such as fields biology, seismology, and genomics. However, Project Pad links the notion of a digital collaboratory with the notion of collaborative commentary. Exploiting spoken word collections provides an opportunity to develop the first such system for extended collaborative commentary.

Listening is an important skill, often assumed to be part of a basic skill set and rarely taught. Good listening -- which means 'critical listening' -- is increasingly valuable in a multimedia-rich world. It should not be assumed and needs to be learned. Project Pad provides the opportunity to teach students these invaluable skills. It incorporates an easy-to-use annotation tool for streaming media. The time-line based tool makes it easy for listeners to identify and attach their own commentary or coding tags to time segments of a stream. Comments are immediately relayed to other participants working on the same work document and are stored permanently for subsequent use and review. An advanced version of the tool permits annotations to the smallest level of detail (e.g., utterances, pauses).

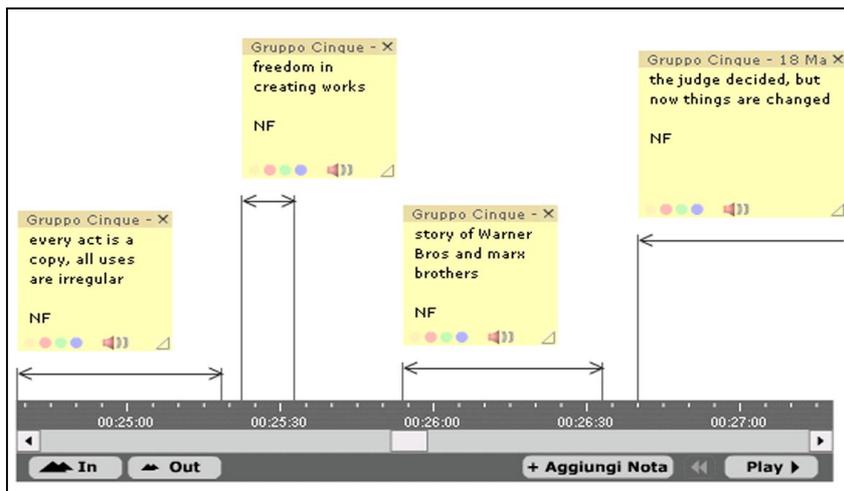


Figure 1: Interface for audio annotation

Fig. 1 shows how the exact length of the audio segment corresponding to the annotation above is delimited by double vertical arrows below the note. Other useful features are: the zoom in/out buttons (to one hundredth of a second or down to the smallest detail of a map, photo or image); the small triangular tab in the bottom right corner of the note which "drags" the annotation to the desired size; the "rewind"/play/pause button for audio or video; the horizontal slide-bar providing random access to the streamed files; the small buttons to change the colour of the notes, which is especially useful for group work. If the audio has a transcription, this can be easily followed while listening thanks to a very effective synchronized highlighting of the aligned text: this feature is of especial value for non-native speakers (fig. 2).

Another invaluable function, added to the Java version of Project Pad in September 2004, is the possibility of highlighting and listening to single phrases or words. Similarly, patterns of sentence stress and intonation can be highlighted and commented: in short, some light can be shed on the "acoustic blur of speech" identified by Richard Cauldwell [8] as the *bête noir* of language learners. No other online application enables the language learner to "zoom in" to word and phrase boundaries in authentic streamed speech and study the effects of elision and assimilation, stress and intonation, those features that make natural spoken English so difficult. On the left-hand side, the interface includes tabs for both individual and group work. By selecting one of the media (audio, video or image) contained in the "Project group" folder, users may exchange comments in real time. By dragging the media object into their own individual folder, users can work independently.

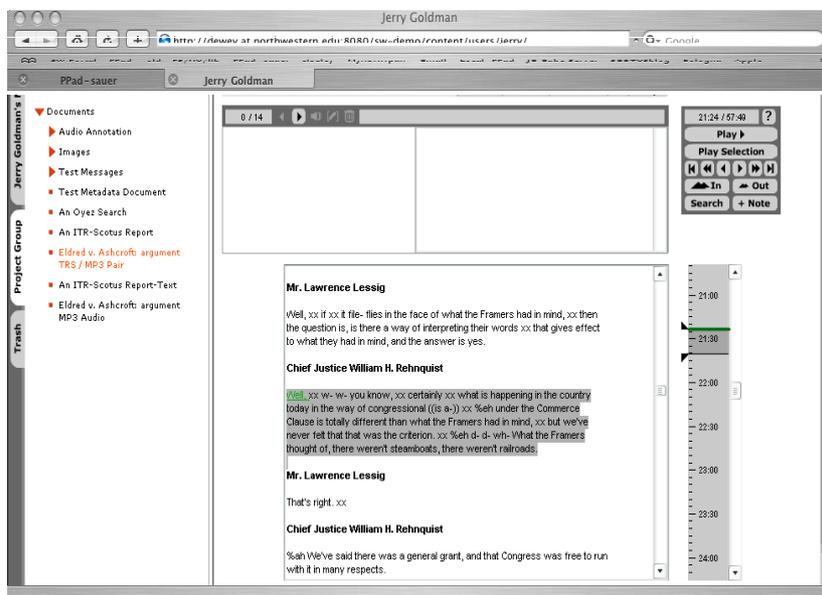


Figure 2: Interface showing text alignment

## 2.2 User-generated data

Project Pad is a Web application written in Java with a user-friendly Flash interface. Close technical collaboration between the group at the University of Bologna and the development staff at Northwestern University, in particular Smith, has enabled the localisation of skills in autonomously elaborating data recorded by local server. This knowledge sharing via Web has been pivotal to the entire project (Gianninoni *et al* [6]). User annotations are stored on the server side in XML format and can therefore be easily converted into "readable" form by applying a style sheet like the one shown in fig. 3. One feature of the style sheet (supplied in input to an XML processor such as Apache Xalan together with the data files) is the conversion of the numerical values referring to the start times of the annotation, its length and the date of insertion.

```
<?xml version='1.0' encoding='iso-8859-1'?>
<xsl:stylesheet version='1.0' xmlns:xsl='http://www.w3.org/1999/XSL/Transform'
  xmlns:java="java" >
<xsl:output method='html' version='1.0' encoding='iso-8859-1' indent='no'/>

<xsl:template match="interval">
<xsl:variable name="dateStr" select="java:util.Date.new()"/>
<xsl:value-of select="java:setTime($dateStr, @modified)"/>
<p>
  <xsl:value-of select="@author-name" />
  (<xsl:value-of select="floor (@start div 60000)" />:<xsl:choose>
  <xsl:when test="((@start mod 60000) div 1000) &lt; 10">0</xsl:when>
  </xsl:choose><xsl:value-of select="(@start mod 60000) div 1000" /> /
  <xsl:value-of select="floor (@duration div 60000)" />:<xsl:choose>
  <xsl:when test="((@duration mod 60000) div 1000) &lt; 10">0</xsl:when>
  </xsl:choose><xsl:value-of select="floor ((@duration mod 60000) div 1000)" />):
  <xsl:value-of select="." /><br/>
  <xsl:value-of select="java:toString($dateStr)"/>
</p>
</xsl:template>
</xsl:stylesheet>
```

Figure 3: XSLT style sheet for *rendering* of XML annotations (by Chris Karr, Northwestern University and Riccardo Gianninoni, CILTA.)

User-generated data of this type could be associated with media in the repository as fields of comments or metadata, thus constituting an important added value. It could also be used to improve the search for specific media by other users. All the annotations added over a specific period of time could be cross-queried, or the same annotations could be applied to a group of similar objects. "It seems logical", Colbath *et al* [9] argue convincingly, "that the end user of the archive should be enlisted in helping to provide the training data. This makes sense since it is likely that the consumer of the data will have the most

familiarity with it, and will be able to provide topics, identify speakers, etc.". Another possible use of the metadata derives from the fact that they are not stored in the repository statically: indeed the potential of Fedora to integrate dynamic transformation mechanisms of stored data enhances its cognitive value. For example, annotations made with Project Pad or the specific transcription tool Transcriber [10] could be transformed on-the-fly into a new medium (QuickTime), which enables the user to listen to the audio and read the synchronised transcription or the comments which have just been added.

### **3 Classroom implementation at the University of Bologna**

The idea of integrating audio materials on the Lockerbie disaster (1988) from the BBC archives into legal English courses largely based on reading and text analysis was inspired by the belief that not only listening, but also reacting to, annotating and commenting the authentic recordings of the voices of lawyers, victims' relatives, forensic experts and journalists involved in that tragic and still 'unresolved' case would increase student involvement and motivation and improve vocabulary retention. In order to measure for these effects, it was first necessary to design and pilot a limited number of activity types and lay the foundation for future data analysis.

#### **3.1 Challenges**

One of the chief problems in implementing Project Pad was the fact that students are already accustomed to perfect Web tools, such as Google, Ebay or PayPal, with their real world outcomes ranging from the exchange of goods to the consummation of a relationship. Motivating students to suspend their disbelief and encourage them to accept responsibility for their role in testing something less than perfect has been a great challenge. Four factors helped us to meet this challenge; personal contact with the Spoken Word principal investigator (Jerry Goldman); the emotional impact and prestige of the audio sources themselves (authentic materials specially requested from the BBC archives, copyright cleared and under special end user agreement with the Spoken Word project at Glasgow Caledonian University); the novelty of being able to do individually paced listening work; and constant contact with no less than three IT staff and instructors. Another interesting challenge has been to encourage students to work on audio material in a non-sequential manner. Foreign language learners are accustomed to one or two listenings in class, after which they are moved on hastily to post-listening activities as described by Cauldwell [12]. In order to learn to work individually with audio files in a non-linear fashion, using the slide bar and zoom function, *ad-hoc* exercise types had to be devised.

#### **3.2 Activity types**

Students (and instructors) needed to carry out a limited number of simple tasks to gain familiarity with the application and to help system administrators and

instructors assess problems to be resolved from session to session with the developers at Northwestern. These included both individual and pair activities; the annotation of a file with simple and compound words; the segmentation of a long file (50 minutes); free choice of a short segment to be annotated in detail individually; pair activities including role play (instructor/student), free commentary, annotation in turns etc. The next pilot sessions in Spring 2005 will introduce other variations, such as the annotation of specific parts of speech or lexical items. Further types of intervention by the tutor, such as the insertion of sentence openers will be tested. All these practice sessions are traced in the data log and by the end of the present academic year will provide a basis for planning statistical analysis with larger student numbers.

### **3.3 Peer collaboration through role-play**

Students were able choose from a number of interactive activities ranging from free annotation to controlled dictation; the most successful however was a "tutor-student" role-play. The following exchange lasted from 13:58:47 to 14:17:51.

User 1: What does the defence want?

User 2: the defence wants the conspiracy charge dropped

User 1: Does the two Libyan defence think that the murder charge will be more difficult to prove than the conspiracy charge? than the conspiracy charge

User 2: yes, the defence hope that the murder charge will be more difficult than the conspiracy charge? than the conspiracy charge

User 1: Is there a fourth judge in the Court?

yes, there will be a fourth judge like in the scottish procedure.

User 2: You can always find a fourth judge in the Scottish procedure.

It is easy to see how a third observer might intervene to correct grammatical or lexical errors; but permanent data storage also ensures that the instructor could address these problems in separate sessions and analyse the exchange purely from a communicative point of view. Interestingly, these two students chose to continue their collaboration along similar question-and-answer lines in a final session, opting to examine a critical passage in a radio report regarding the crown witness whose testimony condemned the accused to a life sentence. The final exchange lasted from 14:35:44 to 15:06:42 and reveals students' awareness of the flimsiness of the evidence for the prosecution. Never before has it been possible to monitor so precisely student exchanges in relation to an authentic data object.

## **4. Future development**

The system we have described undergoes constant change as the working groups plan and implement and as ideas evolve from experience. Harnessing state-of-the-art technology to access and collaboratively interact with prestigious public spoken word collections is indeed ambitious. During the course of the project, new questions will be raised regarding technological solutions for fundamental

aspects, such as the design of new forms of evaluation for listening comprehension abilities, the localisation of the system in other national and disciplinary contexts and the administration of copyright associated with the media (the new format mpeg7 might help to resolve some issues, in that it incorporates metadata into the media itself). The most important factor in the development of an educational, technical and cultural system of this dimension will certainly depend on the ability to communicate and develop online techniques for sharing high-level skills. New figures (such as that of the media specialist) and forms of organisational will become essential in order to manage change in a sector undergoing such rapid transformation.

## References

- [1] EU-US working group on spoken-word audio collections – Executive summary, [www.dli2.nsf.gov/internationalprojects/working\\_group\\_reports/spoken\\_word.html](http://www.dli2.nsf.gov/internationalprojects/working_group_reports/spoken_word.html)
- [2] Spoken Word Project, [www.historicalvoices.org/spokenword/](http://www.historicalvoices.org/spokenword/)
- [3] OYEZ US Supreme Court Multimedia, [www.oyez.org/oyez/frontpage](http://www.oyez.org/oyez/frontpage)
- [4] Lister M., Gianninoni R., Tamburrelli V., Albertazzi S., Brunelli N., The Multimedia Jukebox: Challenges of Implementing a Streaming Server in a University Language Centre, ED-MEDIA, Lugano: AACE, Association for the Advancement of Computing in Education, 2004.
- [5] The Fedora Project, *An Open-Source Digital Repository Management System*, [www.fedora.info](http://www.fedora.info)
- [6] Gianninoni, R., Lister, M., Apprendimento collaborativo sincrono: un sistema aperto (Chapter 6). *E-LEARNING, comunicazione mediata e presenziale*, ed. R.Rossini Favretti, Bononia University Press, Bologna, 2005.
- [7] MacWhinney, B., Martell, C., Schmidt T., Wagner J., Wittenburg P., Brugman H., Broeder D., Hoffert E. Collaborative Commentary: Opening Up Spoken Language Databases, [talkbank.org/projects/lrec-cc.pdf](http://talkbank.org/projects/lrec-cc.pdf)
- [8] Cauldwell, R., Grasping the nettle: The importance of perception work in listening comprehension, [www.speechinaction.com](http://www.speechinaction.com)
- [9] Colbath S., Kubal F., Li D., Srivastava A., Spoken Documents: Creating Searchable Archives from Continuous Audio, *Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000*.
- [10] Transcriber, [www.etca.fr/CTA/gip/Projets/Transcriber/](http://www.etca.fr/CTA/gip/Projets/Transcriber/)