Digital Video Annotations for Education

Authors:
Alexander Nakhimovsky, Colgate University, 306 McGregor, Hamilton NY, 13346, sasha@mail.colgate.edu
Tom Myers, CTO, N-topus Software, Hamilton NY, 13346, tommyers@dreamscape.com

Abstract — We present a Multimedia Annotator that makes it possible to attach multimedia annotations to both text and digital video in the browser window. We illustrate its educational uses, based on several years of experience with a similar desktop program. Our main points are:
• The technological support provided by the Annotator makes it possible for beginning and early intermediate students to use unabridged original films in the target language.
• Such films are better than textbooks for language teaching because they teach language that consists of dialog and is embedded in situation and culture.
• Multimedia annotations attached to a linear narrative sequence are a good format for organizing pedagogical material because they provide a rich set of mnemonics based on episodic narrative memory.
Although we primarily discuss language teaching, the Annotator can also be used to create educational materials for surgeons, craftsmen, journalists, lawyers and other professions.

Index Terms — annotations, commentaries, digital video, HTML, Java, XML, XSLT

MULTIMEDIA APPLICATIONS FOR PEDAGOGY
Multimedia applications -- programs that work with digital video and sound integrated with the more traditional text and graphics materials—are becoming technologically feasible and affordable. Several recent developments have made them so.
• Storage and distribution: The standardization and wide availability of read/write CDROM and DVD drives make it possible to store and distribute large amounts of video data on a single disk or a manageable number of disks.
• Natural language support. The standardization and wide availability of Unicode make it possible to create applications in a variety of natural languages.
• Data formats: The standardization and wide availability of XML data formats make it possible to create cross-platform browser-based applications. At least two major browsers, Mozilla and Internet Explorer, support XSLT, a powerful language for processing and restructuring XML data.
• Browser software: The same two browsers, Mozilla and Internet Explorer, support Javascript DOM interfaces, including on-screen text selection. This makes it possible to highlight and process a region of text corresponding to a video selection.
• Video processing software: Programming interfaces to standard video formats, such as QTJava interfaces for Quicktime, make it easy to control the video from a web server or an applet.

Given these developments, the question of what multimedia can do for education is becoming increasingly pertinent. The answer to that question depends to a great extent on the subject being taught. However, there are also common capabilities that, we believe, should be implemented in any educational uses of multimedia. These capabilities are:
• Video-text synchronization: the ability to associate video data with arbitrary text (not necessarily the sound track of the video), and to establish segment-by-segment correspondence between video and text data.
• Navigation by text and by video: the ability to find a video segment corresponding to a selected stretch of text and vice versa.
• Segment replay: the ability to replay a segment of video while viewing the corresponding text.
• Hypermedia annotation of text and video: the ability to attach arbitrary annotations to a segment of text (and therefore to the corresponding video segment), attach more than one set of annotations to the text, and easily switch between them. (For instance, in language-learning contexts one may want to have grammatical, lexical, stylistic and cultural commentaries.)

Preparing educational materials using the Annotator falls into two stages. In the first stage, the video materia and its accompanying text are broken into a linear sequence of synchronized segments. Each segment can be individually replayed, with the matching text scrolled into view and highlighted. Since the temporal sequence of the video and the spacial extent of text are tightly coupled, one can search the video for a specific text. In addition, the linear sequence of matched text and
video segments creates an infrastructure for video annotations: an annotation to a fragment of text is, in effect, an annotation to the matching video segment.

Once the infrastructure for annotations is set up, the textbook author can begin the really creative part of developing educational materials as a set of annotations on a video-text sequence. An annotation can be any web page, including multimedia materials. More than one set of annotations can be associated with a given video-text pair. Since we use HTML for annotations, the entire complex of educational materials is hyper-linked, with annotations on annotations with links back to the video or anywhere on the Internet.

**MULTIMEDIA LANGUAGE TEACHING PEDAGOGY**

In our own work, we have concentrated on language instruction, traditionally one of the most labor-intensive fields of education. (Our experience in language teaching goes back to the late-1970s, see [1]-[4], [9].) Of all subjects, foreign language instruction can benefit from multimedia materials most obviously and directly. The reason for this stems from the particular qualities of the printed word: compared to movies and other visual materials, the book leaves much more room for the reader’s imagination and background knowledge. In some environments, this feature is a strength. But in foreign language instruction, the call for imagination and background information is a liability: much of what we know is local to a specific time, place and language, and our imagination is frequently driven by deeply internalized local habits of thought.

The narrow bandwidth of a traditional language course, mostly consisting of textual information augmented by illustrations and audiotapes, tends to reinforce the students’ conviction that understanding a foreign language text involves nothing more profound than replacing its words with their dictionary equivalents. Perhaps the most difficult task facing a language instructor is to shake the students out of their linguistic complacency and show them the deep semantic and cultural differences hidden behind dictionary equivalents. Using the movie and its script as the primary “text” of the language course makes that task much easier because in virtually every frame the students can see the clash between their expectations and the realia of a different culture.

To be really effective, movies as language instruction texts should be introduced early in the curriculum. The difficulty with such an early introduction is that students are not yet ready to deal with “unabridged” material created without an expressed pedagogical purpose in mind. On the other hand, pedagogical movies are usually loaded with precisely the cultural assumptions that we want to overcome. As long as the technological means for delivering the movie were the video tape and the laser disk, this difficulty was crippling, and a successful use of the technology was an exception rather than a rule. The advent of digital video makes it possible to alleviate the problem by providing elaborate hypermedia support for the student struggling with the rapid flow of unfamiliar sound from the screen.

**BACKGROUND AND PRINCIPLES**

In the Fall of 1994, the Russian Department at Colgate taught a course whose “reading matter” consisted of digitized excerpts from a Russian movie and the text of that movie, supplemented by an on-screen dictionary and commentary. Students prepared on their own in the lab; they came to class knowing the excerpt inside-out, and ready to discuss it. The course was both a popular and pedagogical success. Both in the classroom and in the lab, students worked well, and their achievement in vocabulary, grammar, and intonation was unusually high. For us as teachers and developers, the course provided valuable lessons in multi-media pedagogy and in the design of a multi-media authoring system. They are discussed in [5]-[7] and briefly summarized below.

From the time of that first course, we have tried to work from general pedagogical principles, largely inherited from our earlier experience as linguists, language teachers and textbook writers, but reinterpreted as guidelines for developing an entirely new technological environment. Principle one, first implemented in our beginning russian textbook and discussed in Nakhimovsky 1984, is that a language course and textbook are more effective if they unfold as a story. We have not done any measurements of performance to test the accuracy of our principle, and we doubt that the results of any such tests could be unambiguous: there are just too many variables to control for. However, our experience, common sense, and what we know about the effectiveness of episodic memory [11] suggest to us that the task of learning is greatly helped if we can say to the student: “Don’t you remember this verb? That’s what Semyon Sergeevich told Katya on their first night out.”

The consequence for a textbook writer is that the presentation of grammatical and lexical material has to be adjusted to the content of the story: the material ends up in a running commentary on the linear text of the story. This is common practice in various “readers” for advanced students but requires a little more effort in an introductory or second-year course. Our experience shows that the results are well worth the extra effort.

Principle two is that materials for language study, whether video or print, work best if they have a richly-textured background full of authentic local detail. In terms of video genres, there is room, of course, for a “talking heads” discussion of deep issues, for a stand-up comedian, or for an animation to illustrate a grammatical point, but the main video for the course and textbook ought to be about people and situations that are perceived as “real” and “different”. It is particularly
important that the people in the story or on screen should be totally removed from the situation of language learning, and should be talking about something that is vitally important to them. It does not matter whether the video is a documentary or a feature film; what matters is that its content not be generated by the grammatical or “situational” needs of language learners. This last point perhaps deserves a bit of explanation.

Particularly in the more difficult languages like Russian, the activities of an intermediate level class are fairly predictable. Students will continue their study of syntax and grammar; they will read some texts, and finally, just as they did in their first-year course, they will engage in role-playing of various sorts designed to give practice in real-life life situations. The role-playing is useful—everyone needs to know how to ask directions. But it is either absurdly artificial (you, Jim, pretend the Kremlin is on your left) or overly focused on student life (what courses are you taking? what’s playing in the local movie house this weekend?). Students in a course like this get good practice in translating the terms of their own lives, and of American popular culture, into the target language. But the foreign culture remains elusive.

By using ordinary movies, produced in the culture, the student-centered, Americanized, or wholly artificial nature of the typical classroom discussion is totally removed. In one of our courses, which uses a romantic comedy, we can discuss the tangled love relations of the characters: these evolve from lesson to lesson, providing a good combination of repetition and novelty. In another course, which uses a film about life in the 1920s, we can talk about the Russian history. All the while, even a short film clip (ours run from a minute and a half to ten minutes) provides a greater range of classroom activities than its corresponding text. Using text alone, we can translate, retell, and role-play. Using a film, we can do all of the above and more: turning the sound off, we can have a student comment on what is happening as it is happening; we can stop the film and describe what is on the table or on the street; we can even supply dialog.

Video material creates new opportunities in teaching grammar as well. For each segment of film, we single out a particular grammatical point for special concentration. Explanations may come from a textbook; drills, as just mentioned, are tied to the movie. Consider verbs of motion, a notorious stumbling block for students of Russian. Motion verbs are usually a feature film; what matters is that its content not be generated by the grammatical or “situational” needs of language learners. This last point perhaps deserves a bit of explanation.

THE ROLE OF THE COMPUTER

Given that our main interest is in the story line as an organizational principle of the course and textbook, it is not surprising that the computer plays a carefully circumscribed role in our system. Specifically, we tried to avoid what we see as three common pitfalls in computer-assisted instruction systems.

First, the computer does not try to do any “teaching”: all the learning takes place between the student and the (video) narrative, not between the student and the computer. The computer’s main role is to deliver the narrative and to help the student cope with it. (See below on the specific ways in which help is provided.) As the preceding section indicates, the students need all the help they can get because we make no concessions in the content of the narrative: it is an unabridged video created without any pedagogical applications in mind.

Second, the computer does not try to “understand” the language or the student. In our project, we eschew all and any applications of artificial intelligence techniques to computer-assisted instruction. Such techniques have been successful only in highly structured domains that are well described by limited sub-languages. In a project like ours, they would consume a good deal of resources without adding significantly to the functionality of the resulting program. Since our goal was to not to investigate a theoretical point but to develop a useful, pedagogically novel and commercially viable application, this kind of cost/benefit analysis proved decisive.

Finally, the computer does not dazzle and distract the student by the interactive and multi-dimensional capabilities of multimedia, keeping the attention focused on the content of the narrative and the linguistic support systems. We sought to minimize the number of buttons on the screen, providing only the navigational controls for moving around in the one-dimensional movie and for studying its text.

USER INTERFACE, ANNOTATIONS, TECHNICAL DETAIL
In a typical lab, the student is assigned a movie excerpt to prepare. The student works through the segments of the movie, looping through a segment as many times as necessary, inspecting the corresponding text, and working with that text using the text support tools. By observing our students in the lab, we found that their actions followed a characteristic pattern. They would begin by watching a segment of the video, perhaps several times. They would then read the corresponding text, invoking support materials (dictionary and commentaries) as necessary. With that done, they would move on to the next segment of the video. In the end, they would watch the entire episode without interruptions. Our interface is designed to support that “action language” in the natural and intuitive way.

All annotation links are standard HTML links and can be created using a text editor or an HTML editor such as DreamWeaver. In general, we have tried to build our system out of standard components, using only standard data formats and APIs (Application Programming Interfaces). Here are the main specifications, most of them discussed in [8]:

- Digital video is in Quicktime.
- Quicktime programming, both for authoring and navigation, is done in Java. It uses QT4J (QuickTime for Java), a Java library from Apple. All Java code is run from a Java applet, loaded at start time.
- Establishing text-video segments is done in Javascript and Java. The resulting XHTML file (i.e., HTML that conforms to the XML standard) is transformed into the HTML of screen display using XSLT (a standard XML transformation language).
- Commentaries and dictionaries are standards-compliant HTML files. Commentary linking and dictionary lookup are done in HTML and Javascript. For larger lexical databases, Java can be used.
- Coordination between selected text and video segments is done in (cross-browser) Javascript.

ACKNOWLEDGEMENTS, ONGOING WORK AND FUTURE PLANS

From its inception, this project has been pursued jointly by Tom Myers, Alice Nakhimovsky and myself. Tom Myers has been the main programmer for the project, and Alice Nakhimovsky has been an early textbook developer and adopter. Slava Paperno of Cornell University has contributed to the project in numerous ways, both by example and specific advice.

The earlier desktop version of the system is in use at Colgate and a few other colleges in the US. We are completing work on a semester-long Russian course, to be used in the fall at Colgate and elsewhere. Work on Chinese and Spanish materials are in early stages. We are actively looking for partners to develop materials for different languages, including specialized languages for business and engineering.

REFERENCES


