Erschienen in: Proceedings of the Language Resource and Evalutation Conference 2004, Lissabon. European Language Resources Association (ELRA), 2004, S. III-VI.

Collaborative Commentary: Opening Up Spoken Language Databases

Brian MacWhinney

Carnegie Mellon University, Psychology
Pittsburgh, PA 15213 USA
macw@cmu.edu

Craig Martell

Naval PostgraduateSchool, Computer Science
Monterey, CA 93943 USA
cmartell@nps.edu

Thomas Schmidt

SFB 538 'Mehrsprachigkeit'
Max Brauer-Allee 60, 22765 Hamburg
thomas.schmidt@uni-hamburg.de

Johannes Wagner

Language and Communication, Southern Denmark University
Campusvej 55, Odense 5230, Denmark
jwa@language.sdu.dk

Peter Wittenburg, Hennie Brugman, Daan Broeder

Max Planck Institute
Wundtlaan 1, Nijmegen, The Netherlands
peter.wittenburg@mpi.nl hennie.brugman@mpi.nl daan.broeder@mpi.nl

Eric Hoffert

Versatility Software
349 Montrose Avenue, South Orange, NJ 07079, USA
eric@versatility-inc.com

Abstract

We define collaborative commentary as the involvement of a research community in the interpretive annotation of electronic records. The goal of this process is the evaluation of competing theoretical claims. The process requires commentators to link their comments and related evidentiary materials to specific segments of either transcripts or electronic media. Here, we examine current work in the construction of technical methods for facilitating collaborative commentary through browser technology. To illustrate the relevance of this approach, we examine seven spoken language database projects that have reached a level of web-based publication that makes them good candidates as targets of collaborative commentary technology. For each database, we show how collaborative commentary can advance the relevant research agendas.

Collaborative Commentary

We define collaborative commentary as the involvement of a research community in the interpretive annotation of electronic records. The goal of this process is the evaluation of competing theoretical claims. The process requires commentators to link their comments and related evidentiary materials to specific segments of either transcripts or electronic media.

Consider the following real-life example. Brian MacWhinney, a researcher in the field of child language, wants to explore evidence for the neo-Vygotskyan claim (Nelson, 1998) that word meanings are shaped through communicative interactions. While browsing through online media at the CHILDES (childes.psy.cmu.edu) site, he locates several instances of videos of mother-child book reading in the Julie and Rollins corpora. In these interactions, mothers help children turn the pages and name the animals or objects in the pictures. In some cases, children call the pictures by the wrong name. In accord with his theoretical views on the logical problem of language acquisition, MacWhinney (in press) believes that mothers will use these errors as opportunities to provide corrective positive feedback. For example, if the child calls a bear a "doggie," the mother should respond, "no, that's a bear, not a doggie." MacWhinney would be

particularly happy if the child subsequently engaged in self-correction and said "a bear." Pursuing this idea, he locates 26 segments in these corpora that are relevant to the position he is advocating.

To initiate the process of collaborative commentary, MacWhinney writes up a short summary of his analysis. He wants to make this analysis available in three ways. First, he wants to post his claim to some discipline-based commentary space on the web. Second, he wants to make sure others who view the relevant segments from the Julie and Rollins corpora are able to see that he has provided detailed interpretive commentary regarding at least 26 specific segments. Finally, he hopes to receive feedback from other researchers regarding his interpretations and arguments.

An Initial Vision

Let us consider for a moment what type of browser display would minimally support MacWhinney's basic goals. Figure 1 presents an initial framework.

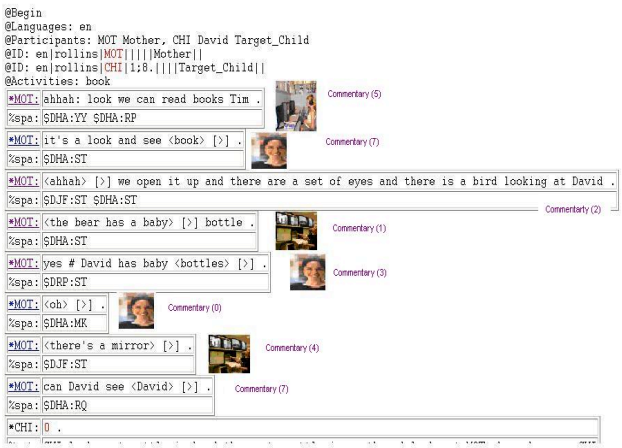

Figure 1: A CHILDES page with commentary links

This figure shows an HTML page displayed dynamically by the CHILDES and TalkBank servers from underlying XML. Thumbnails from the video are used to provide a sense of the nature of the interaction at the relevant point. To the right of the thumbnail is a pointer to the commentary. By clicking on this field, a separate commentary window should open up. This commentary window should allow the research to create the following elements:

- A brief summary of the claim or analysis relevant to the current utterance or utterance sequence.
- Typing of the claims and analysis into specific categories.
- Explanations of the evidentiary role of the texts and media being referenced.
- Links to other texts or claims that are relevant to the current claim.
- Links to external web content, including material (HTML, PDF, Word) that presents the proposed analysis more fully.
- Embedded HTML code.

Once this material has been entered into the commentary field, it can be redisplayed through the current TalkBank SMIL playback facility, as illustrated in Figure 2. In this Figure, the SMIL window echoes the comments that have been entered in the commentary database as continuous streaming playback progresses. It also provides color swatches that code for the nature of the commentary. If a given segment has more attached commentary than can be displayed in the reserved segment of the SMIL window, the window will have a final line listing the number and size of the comments that cannot be displayed.

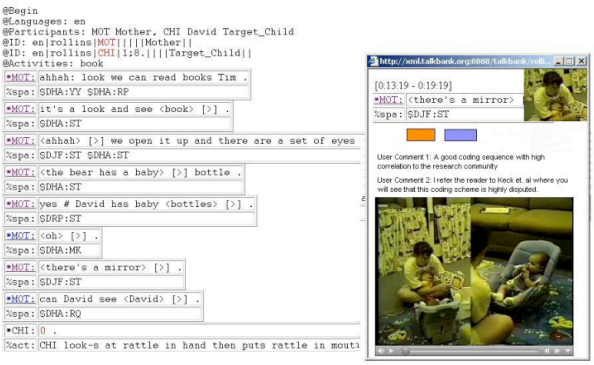

Figure 2: Transcript, SMIL playback, and comments

Possible Implementations

How can we implement this vision? Initially, we might think of turning to "blogging" systems such as Blogger or mBlog. But these systems have no method for linking comments to lines in dynamic web pages that the CHILDES and TalkBank servers produce through on-the-fly generation of HTML from XML. Nor can they support an organized evidentiary database. A more promising alternative is provided by the W3C Annotea project www.w3.org/2001/Annotea that seeks to provide open-source code for building annotation servers. To explore this option, we set up an Annotea server using Zannot inside Zope, accessed by Annozilla inside Mozilla. Although we were able to publish simply commentaries to TalkBank pages using this method, we found that Annotea client development was not keeping pace with new versions of browsers. We realized that it might make sense, of course, to take over the task of Annotea development ourselves. However, before moving in this direction, we wanted to survey other alternatives.

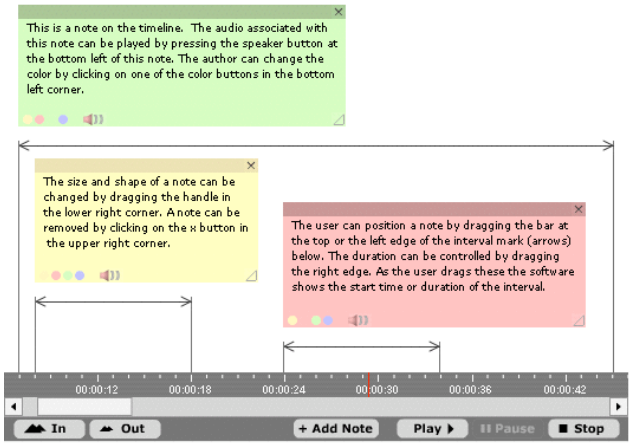

Figure 3: A Project Pad screen

At this point, we learned that Jonathan Smith of the Oyez project at Northwestern University www.oyez.org had built a tool called ProjectPad that showed promise of being able to implement our initial vision. Project Pad (Figure 3) is a program written in Java that controls browsers through Macromedia's Flash. Flash seems particularly well suited for the tight control of commentary entry, item selection, media scrolling, and media linkage that we need to support Collaborative commentary. Moreover, the Java code can also interface well with the Java XML database that will store the commentary in terms of organized evidentiary types. Alternatively it could interface with a SQL database controlled by the AG-API (Ma, Lee, Bird, & Maeda, 2002).

Naked Media

Because TalkBank transcripts are subject to ongoing modifications, reference to line numbers is not stable. A more reliable method links commentary to time points, as in the ProjectPad examples in Figure 3. The idea of linking commentary to media is also in accord with the theoretical emphasis in the Annotation Graph framework of Bird and Liberman (2001). This framework relates all annotations to points in media. The ProjectPad method also opens up a more general possibility for multimedia databases that we will refer to as "naked media." Consider the case of a large database of classroom video data contributed to TalkBank by Rich Lehrer from the geometry lessons of Carmen Curtis. This database consists of 200 hours of classroom video with no accompanying transcripts. It would take perhaps a full year to transcribe all of these sessions. On the other hand, the video can be prepared for streaming web access in about a month. Once the naked video is posted on the web, it can be target for collaborative commentary through ProjectPad. In cases of this type, collaborative commentary can operate effectively even without accompanying transcripts.

Seven Project Areas

To derive a more concrete idea about how collaborative commentary can impact specific research communities, we will survey its application to seven project areas.

Classroom Discourse

Consider the following fictive case of how a researcher in this area will use ProjectPad to produce collaborative commentary. Harriet Keck is a developmental psychologist specializing in children's concepts of number. She and her colleague Robin Clark are both interested in understanding how children solve problems such as 3 + 4 = ?. Keck believes that children solve the problem in an internal mental model and then read out the solution to their fingers. Clark believes that children use their fingers to form external representations of the addends and then count their fingers visually. Keck's model predicts that children will count directly across the fingers, whereas Clark's model suggests that children will begin with placing one addend on each hand separately.

To explore this issue, Dr. Keck uses metadata search tools to find video cases in TalkBank format involving "four-year-old children AND counting". Exploring these videos using the DIVER tool (diver.stanford.edu), she finds that 70% of their gestures support her theory, whereas only 30% are in line with Clark's account.

Over the next several weeks, Keck and her colleagues use ProjectPad to link each case of finger counting to comments that also point to a brief report summarizing her conclusions. Not surprisingly, Clark disagrees with Keck's conclusions and responds by reinterpreting the same video cases that Keck has just analyzed. His analysis points to several counter examples that do not fit Keck's theory. He also argues for including trials that have no overt finger counting in the denominator. Keck, in turn, responds to Clark's criticism by asserting the gestures he has coded are inadvertent hand-movements and revises her paper to anticipate his objection. Keck and students submit their revised paper, including the video data, to the on-line edition of Cognitive Development. One of the reviewers has a question on whether the authors have properly categorized a set of gestures from one of the videos. Keck responds with a close analysis of the gesture in question using a fine-grained analysis of the actual hand-movements. The reviewer is convinced by her response and the paper is published with links to the video data and analysis.

Although this scenario may seem a bit futuristic, it is not very different conceptually from forms of collaborative commentary we have already produced. One example is a special issue of the Journal of the Learning Sciences that focuses on learning about graphs and numerical distributions in a 7th grade classroom. The difference is that in this new framework, analyses will be directly linked to the data, rather than hidden within PDFs. Moreover, in this new framework, analyses will be directly accessible from browsers.

Gesture

Students of gestural communication are becoming increasingly interested in corpus-based research. With appropriate tools, researchers can create corpora coded for grammatical information, discourse structures, and facial expressions, as well as gestures. The resulting corpora can be used to test hypotheses concerning the relationship of the paralinguistic aspects of communication to speech and to meaning. To begin to address this need for a multi-modal corpus, the LDC TalkBank project developed FORM, a non-semantic, geometrically based annotation scheme that allows an annotator to capture the kinematic information in a gesture just from videos of speakers. FORM stores this gestural information using Annotation Graphs (AG), allowing for easy integration of gesture information with other types of communication information. The work so far has produced 30 minutes of FORM-annotated videos of Brian MacWhinney teaching at CMU. This corpus has been published through the LDC and a second one is forthcoming. This corpus and others annotated with FORM can now be opened up to collaborative commentary that either promotes alternative coding systems, revises particular categories, or suggests competing interpretations of the meanings and functions of specific gestures.

Because the FORM corpus has been coded using the AG TableTrans application, it supports the method of storage of collaborative annotations developed by Ma, Lee, Bird, and Maeda (2002). This system allows the user

to store annotations from TableTrans in a central ODBC relational database accessed through SQL queries. The system eliminates problems involved in mapping from AG to SQL by precomputing a table of the transitive closure of annotations. This approach may provide useful methods for structuring and accessing the evidential database needed for a fuller system of collaborative commentary.

Endangered Languages

Approximately 6500 languages are currently spoken worldwide. It can be assumed that around two-thirds of these languages will become extinct in the 21st century. All languages are intimately interlinked with the culture of their speakers, with each representing specific expressions of human thought and social organization. With each language that becomes extinct, priceless intellectual values are lost forever. The DOBES project is dedicated to conserving this cultural heritage. The Max Planck Institute for Psycholinguistics in Nijmegen will house the DOBES data archive and provide user access to the data through a data browser based on the IMDI metadata set. Within this framework, collaborative commentary can function particularly effectively to encourage dialog between field linguists, ethnographers, and anthropological linguists who are interested in analyzing the patterns of symbolic and cultural interaction expressed in recordings of myths, narratives, chants, personal histories, and conversations. These analyses can help us understand the culture-specific and culture-general aspects of communication as reflected in transcripts linked to media. For this commentary to function most effectually, there should be English glosses when possible and links to comparable materials from other languages, both endangered and not endangered.

Conversation Analysis

MOVIN (Microanalysis of Verbal Interaction) is a resource for scholars and students of conversation and interaction. The MOVIN group, with its organizational center at Southern Denmark University, has organized web materials for Conversation Analysis (CA) at conversation-analysis.net The data from this group constitute a major component of current TalkBank materials on conversation analysis, including transcriptions of recorded phone calls from the Nixon Whitehouse, European political television programs, and a variety of classic materials from the CA field. The TalkBank project has been working to link CA format to the TalkBank XML schema in preparation for browsable access and collaborative commentary. Currently, this group organizes frequent face-to-face data analysis sessions of chosen transcripts. It will be interesting to see whether this intensive face-to-face contact can be opened up to wider peer collaborative commentary through the web.

Multilingualism

The work group on Multilingualism at the University of Hamburg has been working for five years to compile a large in-house collection of recordings of bilingual, multilingual, and code-switched interactions. Because this corpus represents data collected in diverse formats in many languages over a period of decades, a primary initial task is the construction of a cross-corpus in-house database. Previous analyses of these data used Partitur-based systems such as SyncWriter that emphasize the linkage of materials to a left-to-right time line. Thomas Schmidt has built the EXMARaLDA editor to provide a method for reformatting these older data and producing newer data in a consistent format with basic XML links to media. To provide for effective collaborative commentary on this rich data set both in-house and outside of Hamburg, we will need to find good ways of supporting Partitur-type display over browser pages. In this regard, ProjectPad seems like a very promising possibility, since it uses the time line as its organizing feature.

Meetings

Another major research area that is currently seeking to engage intensively in collaborative commentary is the field of legal argument analysis. The SCOTUS project at Northwestern is now providing transcripts linked to audio for the past 30 years of oral arguments at the Supreme Court of the United States. Collaborative commentary for this data set will focus initially on a collection of 15 cases in the area of patent law and 15 cases in the area of search-and-seizure. For each area, we will encourage the advocates who argued in the cases to provide commentary on their briefs and answers. We will then organize meetings of legal scholars, commentators, and discourse analysts to provide evaluations of the shapes of the questioning by the justices and the responses of the advocates.

Conclusion

Projects such as TalkBank, LDC, DOBES, MOVIN, DIVER, and SCOTUS have now provided openly accessible databases for the study of spoken language interactions. We now need to implement support for collaborative commentary targeted to these databases. Construction of these new methods will open up many exciting new lines of investigation for each of the several disciplines studying human communication.

References

Bird, S., & Liberman, M. (2001). A formal framework for linguistic annotation. *Speech Communication, 33*, 23-60.

Ma, X., Lee, H., Bird, S., & Maeda, K. (2002). *Model and tools for collaborative annotation.* Paper presented at the Proceedings of the Third International Conference on Language Resources and Evaluation, Paris.

MacWhinney, B. (in press). Multiple solutions to the logical problem of language acquisition. *Journal of Child Language, 29*.

Nelson, K. (1998). *Language in cognitive development: The emergence of the mediated mind*. New York: Cambridge University Press.