Developing Novel Multimodal and Linguistic Annotation Software

Podlasov A., O'Halloran K., Tan S., Smith B., Nagarajan A.
Multimodal Analysis Lab, IDMI, National University of Singapore
9 Prince George's Park, Singapore
{idmpa,k.ohalloran,sabinetan,idmbas,idmarun}@nus.edu.sg

Abstract

In this paper we present a collaborative work between computer and social scientists resulting in the development of annotation software for conducting research and analysis in social semiotics in both multimodal and linguistic aspects. The paper describes the proposed software and discusses how this tool can contribute for development of social semiotic theory and practice.

1 Introduction

Despite the advances that have been achieved in the conceptualization of multimodal theories for analysing language, images and other semiotic resources, many frameworks and models for the transcription and analysis of multimodal data continue to rely on 'low-tech', manual and computer-assisted, technologies. A construct preferred by many researchers for analyzing and representing the interplay of semiotic resources in multimodal data is the *Table*, as exemplified by a range of transcription templates (Baldry & Thibault, 2006). Although tables can be effective for recording which specific semiotic modes and resources are co-deployed at a given moment in the text, they essentially remain bounded by the confines of the printable page.

Besides being laborious to create, page-based analysis severely limits the researchers' ability to effectively capture and portray the complex interplay of semiotic modes and resources as they unfold on a larger scale, especially in the case of dynamic video texts and interactive digital media sites (O'Halloran, 2009). In addition, the analysis will necessarily be restricted to manual transcription and annotation techniques, which often involve a complexity of symbolic notations and

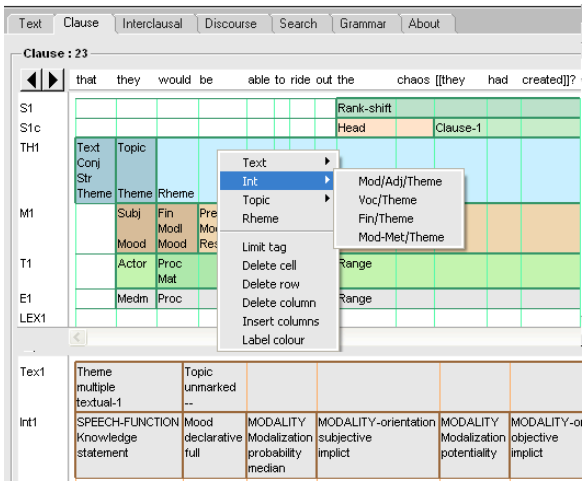

Figure 1. Systemics 1.0 interface.

abbreviations which may be difficult to learn and comprehend (e.g. Thibault, 2000).

We propose that the study of meaning-bearing phenomena requires a collaborative effort between social and computer scientists. Social semiotic study clearly can benefit from theoretical concepts and analytical approaches beyond those traditionally developed for typical social science data, and which draw upon the rich variety of computational techniques developed within computational science. This collaborative approach proved to be productive in development of the *Systemics 1.0* (Judd & O'Halloran, 2002) software for research and teaching *Systemic Functional Linguistics* (SFL) (O'Halloran, 2003). *Systemics 1.0* provides a default systemic functional grammar which can be used to manually code the linguistic analysis from pull-down menus, see Figure 1. The results of the analysis are stored in a data-base format which makes it relatively simple to extract linguistic patterns through relational data base searches. The default grammar can be changed to suit user requirements, so that the software is used for teaching and research purposes. The success of the re-

Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009, pages 134–137,
Suntec, Singapore, 6-7 August 2009.

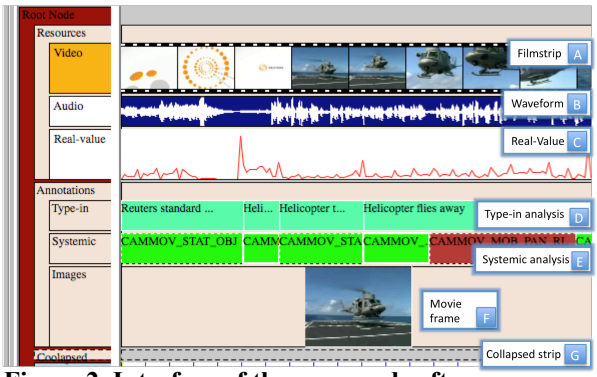

Figure 2. Interface of the proposed software.

search collaboration which resulted in *Systemics 1.0* gives strong motivation to consider computer-based approaches for the more general case of multimodal studies which involves, along with linguistic analysis, analyses of various meaning-bearing phenomena such as gaze, gesture, intonation *etc* (e.g. Bateman, 2008; O'Halloran 2009)

In this paper we report on an ongoing interdisciplinary collaboration between computer scientists and social scientists taking place in Multimodal Analysis Lab, National University of Singapore in the frame of the *Events in the World* project. The project aims at developing an interactive platform for manual, computer-assisted and automatic annotation, analysis and representation of multimedia data for social semiotic and SFL research. The project involves a range of researchers working together to develop the software interface and operational functionalities. The respective roles of the interdisciplinary research team are evolving out of a close collaboration, and the researchers are learning new approaches, perspectives and knowledge which are resulting in new ways of thinking about multimodal phenomena, annotation and analysis.

2 Proposed system

The purpose of the proposed software is to provide social semioticians with the technical means to improve their efficiency and capabilities for studying and annotating multimedia texts, as well as accessing, visualizing and sharing analyses.

2.1 Multimedia resources

Along with supporting standard multimedia resources like video, audio and imagery files, the proposed system supports *real-value resources* to facilitate the use of real-time numerical data in semiotic analysis. Any parameter can be measured during the main media acquisition process or calculated off-line.

2.2 The interface

The interface of the software follows a graphical point-and-click paradigm and is designed to provide easy and straightforward interaction for the user. At the moment of publication, the proposed software is oriented more to the analysis of dynamic time-stamped data, therefore, the main interface window is organized in a partiture-like layout with the horizontal axis corresponding to time. The application's GUI during a sample annotation of a news video clip (© Reuters 2008) is illustrated in Figure 2.

The annotation is organized into strips – containers for semantically grouped annotation elements, which are rendered with respect to the time scale, see Figure 2, D and E. Several strips can be embedded within another strip, which allows hierarchical organization of the annotation data (see Figure 2, 'Annotations' strip with embedded (D) 'Type-in', (E) 'Systemic' and (F) 'Images' strips). A strip can be 'collapsed' (wrapped) in case its display is not needed at any particular time, which is a useful feature when the analysis becomes overloaded with details (see Figure 2, G).

Resources are rendered in the main annotation window in a similar strip-like way. A movie resource control (Figure 2, A) is drawn as a strip of frames enabling interaction with the Movie Viewer tool window. Sound resource control can be drawn as waveform (Figure 2, B). At the moment of publication this control is not interactive and used for display only. Real-value resource control is rendered as a graph (Figure 2, C). Time stamped nodes with associated text or categorical information remain the main tool for annotation and analysis. The node control is rendered as a rectangle and provides the interface for manipulating its position (modifying time-stamps) and displaying the associated data, which is text (Figure 2, D) or categorical association (Figure 2, E) using the Systems Browser tool (Figure 3). The software allows rendering of static image resources. The annotation control representing the image also belongs to a strip in order to follow the general partiture-like organization. The software also allows extraction of frames from the video for annotation in a manner similar to images (see Figure 2, F).

2.3 Categorical analysis

By categorical (systemic) analysis we understand a process of associating time stamped annotation controls (nodes) with choices from a system of

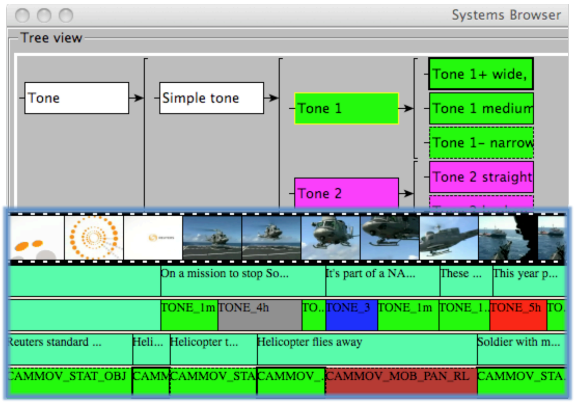

Figure 3. Systems Browser tool window and sample camera movement and tone analysis (superimposed).

categories describing the phenomena semiotically or by other theoretical means. The *Systems Browser* tool window provides the interface to create and manipulate those systems, as well as use them for annotation. Figure 3 illustrates two simple categorical systems: describing camera motion – eg. stationary or mobile camera; the type of tracking (pan, tilt or zoom); and tone choices in speech (cf Halliday 1967). The application allows not only to define the semiotic system itself, but also to associate textual information with every choice, which makes it a helpful interactive glossary for the analyst. Besides that, it allows associating different style attributes of the particular choice: fill color, outline color and dash style, etc. As the user makes systemic choices, the system uses text and style attributes to make the selections visible.

This becomes an important feature of the proposed software. The analysis becomes visible not only textually, but also graphically. In Figure 3 we superimposed the annotation strips with analyzed camera movement and tone for a sample video clip. The filmstrip itself does not visualize the dynamics of the camera in the clip, nor the soundtrack identify tone or other language choices. On the other hand, systemic analysis using style attributes gives a graphical perspective on such choices, so that it is easier to observe and pick up outliers, patterns or points of interest, especially when compared to the traditional page-based annotation. This is especially important when one has to deal with dozen of strips each having hundreds of nodes.

2.4 Overlays

Images and video frames are not dynamic resources and, therefore, are annotated and analyzed by graphical and textual means. Still, textual annotation often needs to be associated with a particular location in the image. Therefore, the proposed software provides a basic graphical annotation tool window – *Overlay Editor*, see Figure 4 - shown also in the thumbnail image in the main interface, giving the analyst the ability to glance over the whole annotation.

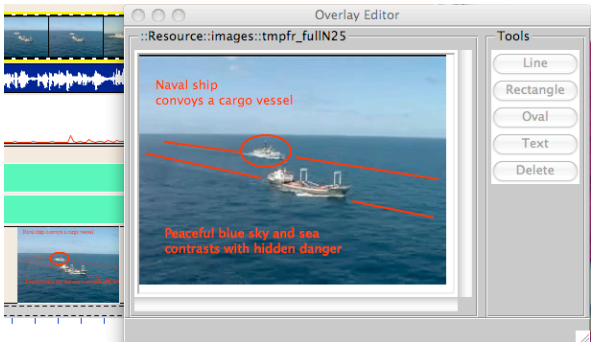

Figure 4. Overlay Editor tool window.

2.5 Visualization

Visualization of the data is especially important in cases when there are multiple analyses of the same text presented or there is a large corpus of texts being analyzed. In terms of the former, visualizations may be useful in picturing correlations, interactions and collaborations between different systems operating concurrently in the multimodal text: *e.g.* linguistic aspects, gaze, gesture, intonation, *etc*. In terms of the latter, visualizations may reveal patterns and departures from patterns within a text. At the moment of publication, the annotation data is visualized by the means of style attributes of systemic choices. However, the architecture allows for more sophisticated visualization techniques to be implemented in future.

2.6 Automation

Generally, the proposed system is designed for the human analyst and, therefore, assumes manual annotation as a default way of producing the analysis data. Nevertheless, the current state-of-art in computer science allows many annotation tasks to be semi- or fully automated. When considering techniques to automate semiotic research, one must keep in mind that the proposed software is designed for use by social semioticians, who study media in many different ways. Therefore, when considering automation techniques we assume that the technique must be general enough to be useful for broad class of applications. For example, one can consider *automatic shot boundary-detection*, *automatic*

face recognition and more general *object tracking* as technologies, which may be employed for automation. Current state-of-the-art algorithms in *automatic speech recognition* may also be considered since many important multimedia genres, like news or studio TV broadcasts, public presentations, and similar, contain clearly spoken speech which is realistic to recognize.

2.7 Templates and collaboration

Different analytical domains – *e.g.* speech and gesture - and theoretical perspectives require different organization of the annotation document and the use of different semiotic systems. The proposed system allows organizing of predefined templates for different analytical purposes. The template is realized as a standalone file where the required structures are defined. The user is provided with a wizard dialog to select the template, which is then imported into the current annotation document.

2.8 Limitations

The scope of proposed functionalities is broad and, therefore, complete coverage is challenging. A main limitation for the proposed software is the ability to automate the analysis. It is very unrealistic to consider that human analyst can be removed from the process of annotation, since the domain itself, which is social semiotics, is fundamentally human-oriented. Therefore, we consider that semiotic analysis will remain manual to a large extent, and the automation will be employed mostly for routine and technically implementable tasks.

3 Conclusions

In this paper we present a collaborative work between computer and social scientists developing a software platform facilitating research in social semiotics. The proposed software is used for annotation and analysis of multimodal data supporting different media types simultaneously (video, sound, images, text and real-value data) and it provides intuitive GUI for entering the analysis. It supports innovative categorized (systemic) analysis and provides the library of predefined analytical systems developed in the literature together with the ability to create and customize new systems. In addition, the software provides the interface for graphical analysis of imagery data and provides functions for data visualization.

The project contributes to social semiotics by providing a specialized domain-oriented software tool which significantly increases the productivity of the analyst resulting in more extensive studies and better-grounded theories. The proposed visualization interface makes it easier to see data patterns, outliers and recognize points of interest. The ability to customize descriptive systems facilitates experimentation and promotes the development of theoretical frameworks in social semiotics. Unified frameworks for annotation of different media types encourage cross-domain analysis and integration of research from different fields. Though the project is still at the development stage, the obtained results and feedback from practicing social semioticians are promising.

Acknowledgments

This work was supported by the Singapore National Research Foundation (NRF) Interactive Digital Media R&D Program, under research Grant NRF2007IDM-IDM002-066 awarded by the Media Development Authority (MDA) of Singapore.

References

A. Baldry, P. J. Thibault, Multimodal Transcription and Text Analysis, Equinox, London, 2006.

Tan, S. 2005. A Systemic Functional Approach to the Analysis of Corporate Television Advertisements. MA Thesis, Department of English Language and Literature, National University of Singapore.

P. J. Thibault, "The Multimodal Transcription of a Television Advertisement: Theory and Practice", in Anthony Baldry (ed.) Multimodality and Multimediality in the Distance Learning Age, pp. 311-385, Palladino Editore, Campobasso, Italy, 2000.

K. L. O'Halloran, "Multimodal Analysis and Digital Technology", In A. Baldry and E. Montagna (eds.), Interdisciplinary Perspectives on Multimodality: Theory and Practice, Proceedings of the Third International Conference on Multimodality, Palladino, Campobasso, 2009.

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. The Hague; Paris: Mouton.

K. L. O'Halloran, "Systemics 1.0: Software for Research and Teaching Systemic Functional Linguistics", RELC Journal, vol. 34(2), pp. 157-178, 2003.

Bateman, J. (2008). Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents. Hampshire: Palgrave Macmillan.

Judd, K. & O'Halloran, K. L. (2002) *Systemics 1.0.* Singapore: Singapore University Press.