Article

Journal of Mixed Methods Research
2018, Vol. 12(1) 11–30
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1558689816651015
journals.sagepub.com/home/mmr
SAGE

A Digital Mixed Methods Research Design: Integrating Multimodal Analysis With Data Mining and Information Visualization for Big Data Analytics

Kay L. O'Halloran[1], Sabine Tan[1], Duc-Son Pham[1], John Bateman[2], and Andrew Vande Moere[3]

Abstract

This article demonstrates how a digital environment offers new opportunities for transforming qualitative data into quantitative data in order to use data mining and information visualization for mixed methods research. The digital approach to mixed methods research is illustrated by a framework which combines qualitative methods of multimodal discourse analysis with quantitative methods of data mining and information visualization in a multilevel, contextual model that will result in an integrated, theoretically well-founded, and empirically evaluated technology for analyzing large data sets of multimodal texts. The framework is applicable to situations in which critical information needs to be extracted from geotagged public data: for example, in crisis informatics, where public reports of extreme events provide valuable data sources for disaster management.

Keywords

multimodal discourse analysis, social semiotics, data mining, information visualization, digital mixed methods design

Mixed methods research is defined as:

> [a]n approach to research in the social, behavioural, and health sciences in which the investigator gathers both quantitative (closed-ended) and qualitative (open-ended) data, integrates the two, and then draws interpretations based on the combined strengths of both sets of data to [better] understand research problems. (Creswell, 2015, p. 2)

[1]Curtin University, Perth, Western Australia, Australia
[2]Bremen University, Bremen, Germany
[3]KU Leuven, Leuven, Belgium

Corresponding Author:
Kay L. O'Halloran, Curtin Institute for Computation and School of Education, Curtin University, Kent Street, Bentley, Perth, Western Australia 6102, Australia.
Email: kay.ohalloran@curtin.edu.au

In this article, a new mixed methods research design is presented where qualitative data are transformed into quantitative data in a digital environment, to use data-mining and information visualization techniques for analysis of large data sets. The digital approach to mixed methods research is illustrated through a research framework which combines qualitative methods of multimodal discourse analysis (i.e., the analysis of language, images, and other resources) with quantitative methods of data mining and information visualization in a multilevel, contextual model for analyzing large data sets of multimodal texts involving language and images.

In order to situate the proposed digital approach to mixed methods research, existing research designs are first reviewed. Following this, the digital approach is illustrated through a research framework for integrating multimodal discourse analysis with data mining and information visualization (henceforth, referred to as the ''multimodal mixed methods research framework,'' or more simply, ''the multimodal research framework''). The theoretical foundations and existing computational methods for multimodal discourse analysis are then reviewed before discussing the multimodal research framework in detail.

Mixed Methods Research Approaches and Design

There are several basic research designs in mixed methods research (e.g., Creswell, 2015; Creswell & Plano Clark, 2011; Curry & Nunez-Smith, 2015; Morse & Niehaus, 2009). For example, Creswell (2015) proposes three basic designs: (a) ''convergent designs,'' where qualitative and quantitative data are merged and interpreted; (b) ''explanatory sequential designs,'' where the analysis and interpretation of quantitative data are followed by a qualitative analysis to explain the quantitative results; and (c) ''explorative sequential designs,'' where the problem is first explored through qualitative data collection and analysis, followed by a quantitative phase to develop and apply an instrument design or intervention. The impact of the application of the instrument and the intervention are then explored quantitatively.

Advanced mixed methods designs typically use one of these basic research designs, but add new dimensions and features. For example, Creswell (2015) describes three advanced designs (see also Guetterman, Fetters, & Creswell, 2015):

- *Intervention Design.* A problem is studied by conducting an experiment or intervention and qualitative data are added before and afterward to understand the problem and to explain the outcomes.
- *Social Justice Design.* The basic mixed methods design is studied within the context of a social justice lens; for example, gender, race, social class, disability, lifestyle, or combinations of these lenses.
- *Multistage Evaluation Design.* A study is conducted over time to evaluate the success of a program or activities by developing and testing quantitative measures, implementing the program, and conducting follow-ups.

In these basic and advanced mixed methods designs, qualitative and quantitative data often remain as distinct data sets which are used for various purposes such as needs assessment, theory conceptualization, instrument development, implementation and testing, and program follow-up and refinement. As such, the meaningful integration of qualitative and quantitative data sets is often not achieved, despite the fact that this is a desired goal of mixed methods research (Fetters, Curry, & Creswell, 2013; Guetterman et al., 2015). In some cases, qualitative and quantitative data sets are merged using computational techniques; for example, qualitative coding is transformed into numerical values for statistical analysis (e.g., frequency counts, measures of association, etc.) and the results are represented as visual displays of relationships, most often as multidimensional scaling, cluster analysis, and correspondence analysis (see, e.g., Bazeley, 2010). Building on these approaches, Guetterman et al. (2015) explore joint displays as a way of bringing qualitative and quantitative data together visually to discuss the integrated analysis and draw out new insights. For example, joint displays can show ''statistics-by-themes'' and side-by-side comparisons, and connect research findings to theoretical frameworks and recommendations.

In the proposed digital approach, data merging and integration are expanded to include the transformation of qualitative data into quantitative data for the development and use of data-mining and information visualization techniques for mixed methods research. In this new design, patterns and trends are explored and compared along different data dimensions in an interactive digital environment, bringing forth insights and permitting testing of hypotheses in an iterative and interactive way. The research design, called the *digital mixed methods design*, is based on a multistage evaluation design, but, in this case, the quantitative results derived from the qualitative analysis provide the basis for developing data mining and information visualization as analytical tools for mixed methods research. The digital mixed methods design involves:

> Qualitative analysis → Transformation to quantitative data → Data mining → Information visualization → Qualitative exploration

For example, as discussed below, qualitative analyses of text, images, and other resources are transformed into quantitative data using digital tools, in this case, multimodal annotation software. These quantitative results inform machine-learning techniques which are applied for data mining and information visualization of large multimodal data sets in a multilevel contextual model, in order to feed back into qualitative interpretations and explanations of large data sets. As such, qualitative and quantitative methods are integrated in relation to the design of the study, the methods, interpretation, and reporting (Fetters et al., 2013), in this case providing new opportunities for advancing the science of mixed methods research.

The digital mixed methods design is illustrated through the multimodal research framework, presented here with respect to the requirements of crisis informatics, an emergent, interdisciplinary field of study that arose in response to major disasters such as the Asian Tsunami in 2004 and Hurricane Katrina in 2005: namely, how to extract useful information from public reports posted online during times of crisis (Pipek, Liu, & Kerne, 2014). The quality of collaboration between governmental, professional, volunteer, and citizen responders has a major impact with respect to loss of lives and property in crisis situations (Pipek et al., 2014). However, current tools for extracting meaning from public reports of extreme events in disaster zones fail to provide deep interpretive insights into how people are reacting to crisis situations, largely because existing models of communicative behavior fail to relate the message (i.e., the actual words and images) to context and meaning. In the multimodal research framework, this problem is overcome using the theoretical foundations of social semiotic approaches to multimodal (discourse) analysis combined with contextual information about human social life derived from Wikipedia, as discussed in the following sections (for further illustration of the framework design and methodology, refer to Figure 6).

The multimodal research framework can be applied to any mixed methods study where text, images, video, and other multimodal data are analyzed. More generally, the digital mixed methods design has implications for the future directions of mixed methods research.

Theoretical Foundations of Multimodal Discourse Analysis: A Social Semiotic Approach

Despite the abundance of digital data, detailed empirical analysis and visualization of human discourses remain an elusive but valued goal. For example, while Twitter and other social media have proved useful for ''dissemination, information gathering, and as inputs to situational assessment'' (MacEachren et al., 2011, p. 183), most geovisual tools are concerned with tracing the spatiotemporal processes of information diffusion to answer questions such as ''*when, where, and how an idea is dispersed*'' (Cao et al., 2012, p. 2650), rather than tracking the meaning which is communicated.

In order to track meaning, methods are required for making it accessible to codification, quantification, and characterization in terms which are appropriate for subsequent theorization and practical application. This in itself is a major challenge and one of the main barriers to progress. In recent decades, however, multimodal discourse analysis has emerged as an interdisciplinary area of research providing powerful analytic frameworks precisely targeting the meanings that are exchanged in social interactions. This field, arising out of linguistics and semiotics and drawing on other relevant fields (e.g., film studies, design, cognitive sciences), involves the study of the contributions and interactions of linguistic and nonlinguistic modes (e.g., spoken and written language, image, gesture, sound, page layout, and website design) in the communication of meaning (Jewitt, 2014).

The approach to multimodal analysis adopted here is multimodal social semiotic theory, a theory of meaning in which semiotic resources (e.g., words, images, and sounds) are conceptualized as interrelated systems which together constitute and manifest culture (Halliday, 1978; Halliday & Hasan, 1985). Semiotic resources are defined as having a meaning potential which is captured in terms of interconnected systems of meaning. Following Halliday's systemic functional theory, such systems are organized according to the functions (called ''metafunctions'') which the resources serve in society: (a) *experiential and logical meaning*: to structure experience of the world; (b) *interpersonal meaning*: to enact social relations and create a stance toward happenings and entities in the world; (c) *textual meaning*: to organize experiential, logical, and interpersonal meanings into messages (Halliday & Matthiessen, 2014; Martin & Rose, 2007). The messages exchanged in any communication system are characterized in terms of options selected from these systems. For example, systems are formulated for language (Halliday & Matthiessen, 2004), static images (Kress & van Leeuwen, 2006; O'Toole, 2011), music (van Leeuwen, 1999), and film resources (Bateman, 2014a). Systems are typically organized according to different ranks of constituency (e.g., discourse and clause for language; and work, episode, and figure for image), as illustrated in the examples of text and image systems given in Table 1. These are based on Halliday's (Halliday & Matthiessen, 2014) and Martin's (Martin & Rose, 2007) systems for language and O'Toole's (2011) framework for images. In many respects, this is analogous to approaches within visual studies and communication studies that attempt to provide standardized (and hence quantifiable) methods for qualitative visual analysis (e.g., see Margolis & Pauwels, 2011); here, however, the categories adopted have been related to broader schemes of functional interpretation both within and across modalities and artifacts.

For example, the pictorial framework in Table 1 can be used to analyze images, such as the one in Figure 1, which is a photograph provided by the Federal Emergency Management Agency in the U.S. Department of Homeland Security.[1] The photograph shows some members of the Colorado Task Force 1 (CO-TF1) at the World Trade Center following the terrorist attacks on September 9, 2001. The distinctions in Table 1, although not exhaustive, provide a robust scaffold for drawing out of the image the interpretations that viewers will typically make, at the same time making explicit qualities of the image that support those interpretations, as explained below.

In terms of the composition, there is a relative placement of figures in the foreground with respect to a background. Compositional Vectors (vertical) are formed by the three figures in the foreground and the remaining parts of the buildings in the background, contrasting with at best oblique vectors among the debris. In terms of experiential meaning, the figures are clearly

Table 1. Examples of Text and Image Systems.

Semiotic resource/metafunction/rank	System	Description
Text		
EXPERIENTIAL		
Clause	Processes; Participant Roles; Circumstance	Happenings, actions, and relations
INTERPERSONAL		
Clause	Speech Function	Exchange of information (e.g., statements and questions) and goods and services (e.g., commands and offers)
TEXTUAL		
Clause Discourse Semantics	Information Focus	Organization of information, with points of departure for what follows
Pictures		
EXPERIENTIAL		
Work	Narrative Theme; Representation; Setting	Nature of the scene
Episode	Processes; Participant Roles; and Circumstance	Visual happenings, actions, and relations
Figure	Posture; Dress	Characteristics of the participants
INTERPERSONAL		
Work	Angle; Camera Distance; Lighting	Visual effects
Episode	Proportion in Relation to the Whole Image: Focus; Perspective	Happenings, actions, and relations with respect to the whole image
Figure	Gaze—Visual Address	Direction of participant's gaze as internal to image or external to viewer
TEXTUAL		
Work	Compositional Vectors; Framing	The organization of the parts as a whole, with the visual marking (e.g., framing) of certain parts
Episode	Relative Placement of Episode; Framing	Position of the happenings, actions, and relations in relation to the whole image, and the visual marking of certain aspects
Figure	Relative Placement of the Figure Within the Episode; Arrangement; Framing	Position of figures in relation to happenings, actions, or relations, and the visual marking of certain aspects of those figures

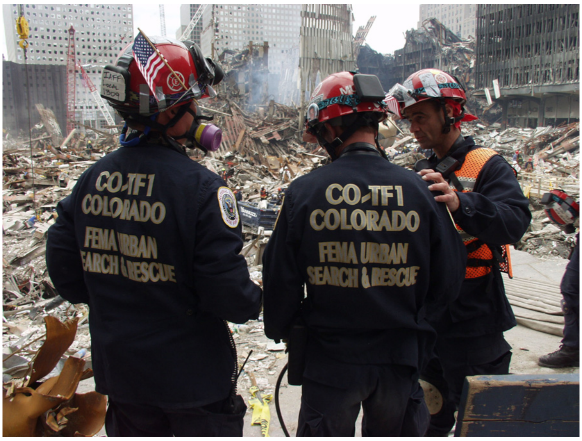


Figure 1. Colorado Task Force I (CO-TFI) at the World Trade Center, September 2001, Federal Emergency Management Agency in the U.S. Department of Homeland Security.[1]

identifiable as members of the CO-TF1, given the choices made for Dress (uniform with textual identifiers). The interpersonal meaning then makes a particularly significant contribution: Even though the Task Force member on the right is smaller and partially obscured by the larger middle figure, the access granted to his facial expression and gaze makes him critical to the interpretation. As such, the figure on the right assumes a central Participant Role (as an agent in behavioral processes), realized through facial expression, gaze, stance, and gesture. Further interpersonal meanings are contributed by the choices for the systems of Angle (approximately eye level), Camera Distance (medium-close), Lighting (naturalistic), Proportion in Relation to the Whole Image (naturalistic), Focus (in focus), Perspective (foreground), and the Gaze—Visual Address (internal to the image itself). These function to position the viewer as a close observer of the scene. Combining the meanings described, we have a reading for the Work as a whole, consisting of a Narrative Theme stemming from the central episode where one figure (right) appears to be consoling the two other figures (center and left) amid the devastation caused by the attacks. The analysis thus reveals how the meanings made through the combination of choices from the visual systems displayed in Table 1 work to construct an experiential account of an extreme event, in this case, the World Trade Center attacks in New York, and at the same time, contribute to create an interpersonal stance toward those events. The analysis of a series of multimodal texts related to an extreme event (e.g., such as an act of violent extremism), could thus offer valuable insights into potential changes in the content and tone of public discourse over space and time, for example, shifting from a focus on victims, perpetrators, rescue workers, to the wider social implications of the event itself.

Choices from multimodal systems of meaning form more or less stable (but evolving) configurations, which are socially and culturally recognizable. These configurations are described through register theory (Halliday, 2002; Matthiessen, 2009), which posits that discourse in particular social contexts—for example, academic writing (Gardner, 2012a, 2012b), health care (Lukin, Moore, Herke, Wegener, & Wu, 2011; Matthiessen, 2013), and translation studies (Hatim & Mason, 1990; Steiner & Yallop, 2001)—form recognizable patterns. That is, while the meaning potentials of language and images are diverse, the actual options selected in any

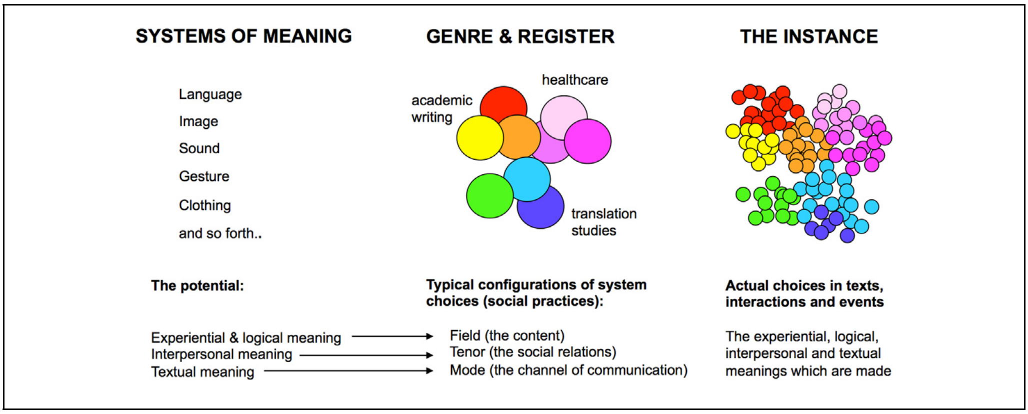

SYSTEMS OF MEANING
GENRE & REGISTER
THE INSTANCE
Language
Image
Sound
Gesture
Clothing
and so forth..
healthcare
academic writing
translation studies
The potential:
Typical configurations of system choices (social practices):
Actual choices in texts, interactions and events
Experiential & logical meaning
Interpersonal meaning
Textual meaning
Field (the content)
Tenor (the social relations)
Mode (the channel of communication)
The experiential, logical, interpersonal and textual meanings which are made

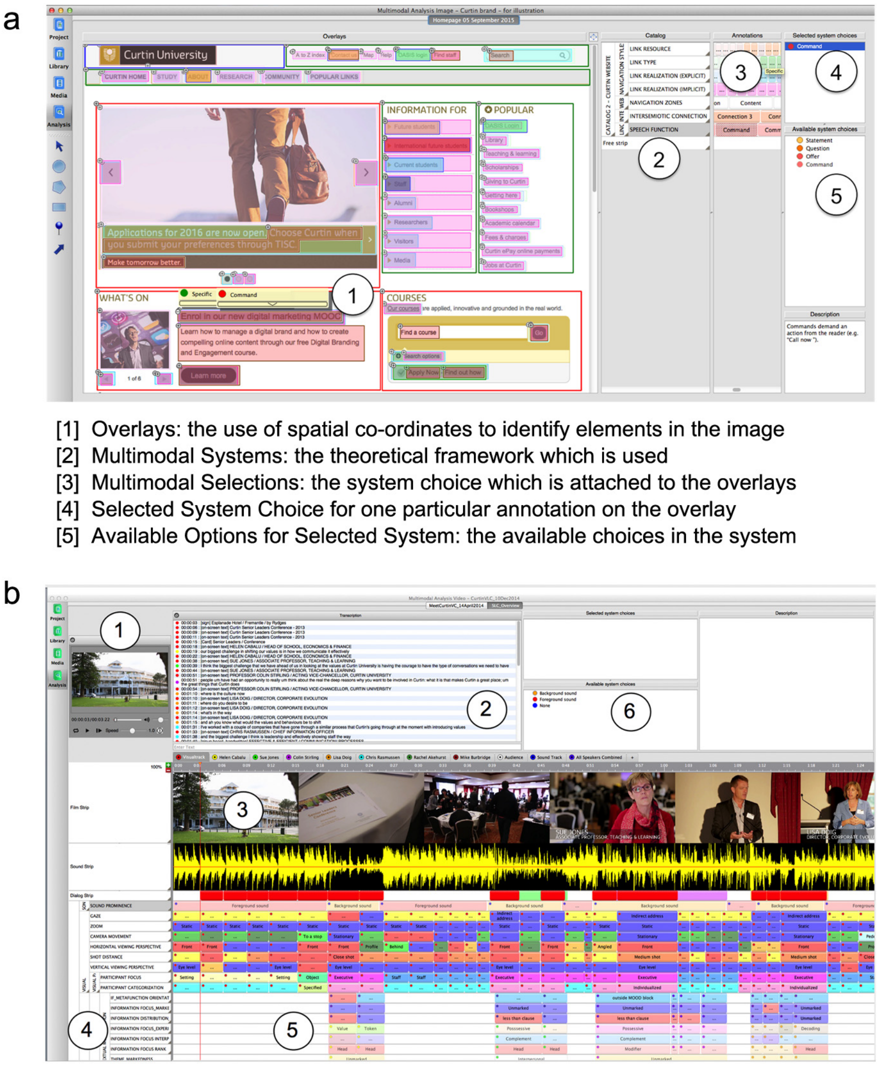
Figure 2. The multimodal social semiotic model.

context are preconditioned by previous configurations of choices within that culture. In this regard, multimodal choices are constrained contextually according to three key dimensions: field—the nature of the social activity (realized through experiential and logical choices); tenor—the social relations which are enacted (realized through interpersonal choices); and the mode—spoken, written, and visual forms of representation (realized through textual choices), as displayed in Figure 2. In other words, social practices are patterned, making it possible to identify key system choices in operation in any context, making the multimodal social approach applicable to any mixed methods study involving the analysis of texts, images, videos, and other media (e.g., social media).

Computational Methods and Tools for Multimodal Discourse Analysis

While multimodal discourse analysis provides a rich array of theoretical tools, the latest challenge in the field has been the move from manual analysis and discursive interpretation of a limited number of multimodal texts toward automated recognition of multimodal meanings across large data sets (Bateman, 2014a; O'Halloran, Chua, & Podlasov, 2014) and corpus-based empirical grounding and testing of insights (Bateman, 2016). To handle the complexity of such analyses, software applications (e.g., NVivo,[2] Atlas.ti,[3] and ELAN[4]) with facilities for (a) importing media files (text, images, and/or videos), (b) developing paradigmatic frameworks of analysis, (c) annotating the text and media files, and (d) storing and exporting the analyses for further data processing can be used (see multimodal annotation tools in Bateman, 2014b).

In this case, *Multimodal Analysis Image*[5] and *Multimodal Analysis* Video,[6] spin-off technologies developed by Kay O'Halloran and colleagues in the Multimodal Analysis Lab in the Interactive & Digital Media Institute at the National University of Singapore, will be used since these applications have been designed precisely to explore register-based multimodal patternings of the kind we target. The software applications permit media files to be imported and analyzed using different sets of analytical categories (cf. the systems set out in Table 1), represented as system networks of options (e.g., the system of ''Gaze—Visual Address'' has the options of ''Direct,'' ''Indirect,'' and ''None''). The results of the analyses are stored in a database for further data processing and visualization. Catalogues of system networks can be freely defined and used in the software applications, so the platforms are customizable for different contexts and research goals (e.g., O'Halloran, E, & Tan, 2015; O'Halloran, Wignell, &


a
Multimodal Analysis Image – Curtin brand – for illustration
1
2
3
4
5
b
Multimodal Analysis Video – CurtinVLC_10Dec2014
1
2
3
4
5
6

[1] Overlays: the use of spatial co-ordinates to identify elements in the image
[2] Multimodal Systems: the theoretical framework which is used
[3] Multimodal Selections: the system choice which is attached to the overlays
[4] Selected System Choice for one particular annotation on the overlay
[5] Available Options for Selected System: the available choices in the system

[1] Video Player: facilities for playing, pausing, slowing down etc. the video
[2] Transcription Tool: facilities for transcribing the video according to speaker
[3] Film Strip: the visual display of the video frames over time
[4] Multimodal Systems: the theoretical framework which is used
[5] Time-stamped Annotations: the coding of system choices in nodes which have time-stamps attached to them
[6] Available Options for Selected System: the available choices in the system

Figure 3. Multimodal analysis: (a) Multimodal analysis of text and images and (b) Multimodal analysis of a promotional video.

Tan, 2015). The benefits of using the purpose-built software applications, combined with computational methods, are illustrated further below.

The analyses of the two multimodal texts displayed in Figure 3, which shows a screenshot of a university website[7] (Figure 3a) and a university promotional video[8] (Figure 3b), have been

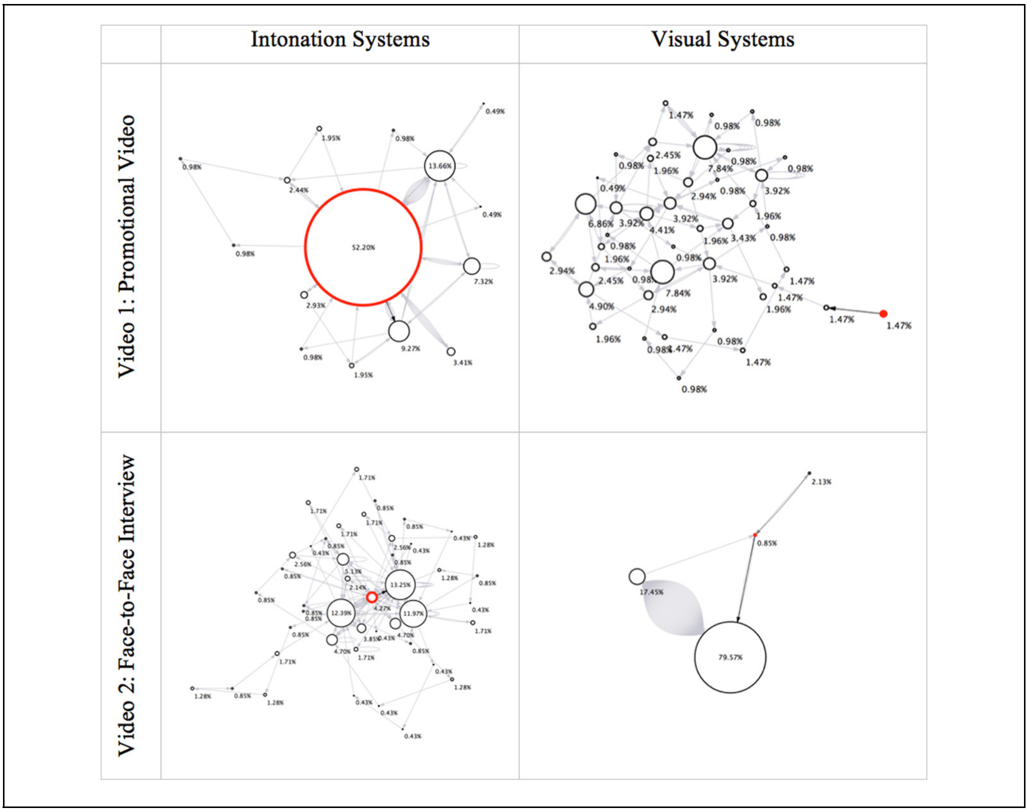


Figure 4. Visualization of intonation and visual systems utilized in different video genres (Tan et al., 2015).

undertaken using *Multimodal Analysis Image* and *Multimodal Analysis Video* software (Tan, Smith, & O'Halloran, 2015). In Figure 3a, the website has been manually annotated by creating overlays (in the form of rectangles, circles, polygons, lines, or pins) on the image and assigning choices from various system networks (i.e., which contain various options) to those overlays. In Figure 3b, the video has been manually transcribed, and the video is analyzed according to a range of multimodal systems by creating time-stamped annotations and assigning system choices to those nodes from the available options. In each case, the annotations of the system choices are stored in a database for later retrieval and visualization of the data.

The analytical procedure by which the results of the multimodal analysis of text, images, and videos are stored in a database means that qualitative analysis is transformed into quantitative data. These data are directly related to the media (in terms of spatial coordinates for text and image files and time stamps for video files), permitting patterns and trends to be revealed. For example, analyses of the Curtin University website, as depicted in Figure 3a, reveal that brand communications for students have shifted over time from a narrow focus on certain domains of experiential meaning (e.g., courses, regulations) to domains that include lifestyle, the student experience, employment, and future prospects in an interactive environment (Tan et al., 2015).

In addition, the results of the multimodal analysis can be visualized to discern discourse patterns. For example, the state transition diagrams in Figure 4 show combinations of semiotic choices, displayed as ''states'' (i.e., the circles), and the movements between those combinations, displayed as ''transitions'' (i.e., the lines) in dynamic media (e.g., videos). Each state records the choices made from the systems defined for the semiotic resources for the selected units of analysis; the transitions then document how such combinations of semiotic choices change as a text unfolds. The state transition diagrams in Figure 4 reveal how semiotic choices integrate and unfold dynamically in the promotional video displayed in Figure 3b (see Figure 4: Video 1, top). Such patterns can then be compared with other contexts, for example, a recorded face-to-face interview at the same university (see Figure 4: Video 2, bottom; Tan et al., 2015).

In Figure 4, the combinations of linguistic and visual choices (i.e., the states, depicted by the circles), displayed as relative percentages of the total video time, together with the movements between those states (i.e., the transitions, depicted by the lines) reveal that visual systems (e.g., camera angle, camera movement, and shot distance) are actively exploited in the promotional video (Video 1, top right), compared with the intonation systems (e.g., tone, information focus, and information markedness) which are less varied in their use (Video 1, top left). The converse is true for the face-to-face interview, where visual systems are deployed in a more routine and restricted fashion (Video 2, bottom right) compared with the intonation systems which are more actively utilized (Video 2, bottom left). In this case, the varied use of intonation functions to create a ''collegial'' style of leadership in the face-to-face interview, while the use of visual systems and other discourse patterns create a more ''managerial'' leadership style in the promotional video (Tan et al., 2015). Such semantic patterns, emerging from computational processing of manual multimodal analysis, are shown to ''resonate with the aims of the producers of the different media'' (O'Halloran, E, & Tan, 2015, p. 408), providing a methodology for building profiles of key system choices which are called into play within and across different media and contexts.

In summary, the transformation of qualitative data into quantitative data in a digital environment has significant outcomes, which include the display of data as interactive visualizations for interpreting the texts and developing profiles of key systems in different contexts. The resulting multimodal data, stored and retrieved from such platforms, can then be modeled and further analyzed using a range of mathematical and computational techniques, as found in other mixed methods designs. For example, singular value decomposition, *k*-means clustering and temporal logic have been used to detect semantic patterns in multimodal texts which would not otherwise be discernible; for example, O'Halloran, E, Podlasov, and Tan (2013) and O'Halloran, E, and Tan (2014) characterize in detail the differences in the multimodal strategies deployed by a climate scientist and a climate denialist and how they are visually portrayed in a televised debate about climate change, illustrating how the approach can be used for other studies.

These approaches described above have nevertheless several major drawbacks that are addressed in the research framework we now present. First, it is not possible to model and predict discourse patterns extrapolating from a limited number of detailed analyses. Second, the modeling of multimodal data using dimensionality reduction and clustering techniques results in visual patterns that require a human analyst to make sense of them, rather than delivering explicit, computable accounts of the semantic patterns which have been derived. Whereas some proposals suggest that this is sufficient and that the human propensity to recognize patterns will be able to do the required work (Manovich, 2012), this is unlikely to be a successful strategy for large-scale data and complex and targeted investigations, where more focused and formalized methods are essential for understanding semantic patterns in complex data sets.

In the multimodal mixed methods research framework set out here, multimodal analysis is used to identify key semantic patterns in text, images, and videos. From here, machine learning based on these semantic patterns is used to develop data-mining techniques for the analysis of larger data sets. These data-mining techniques include, for example, natural language

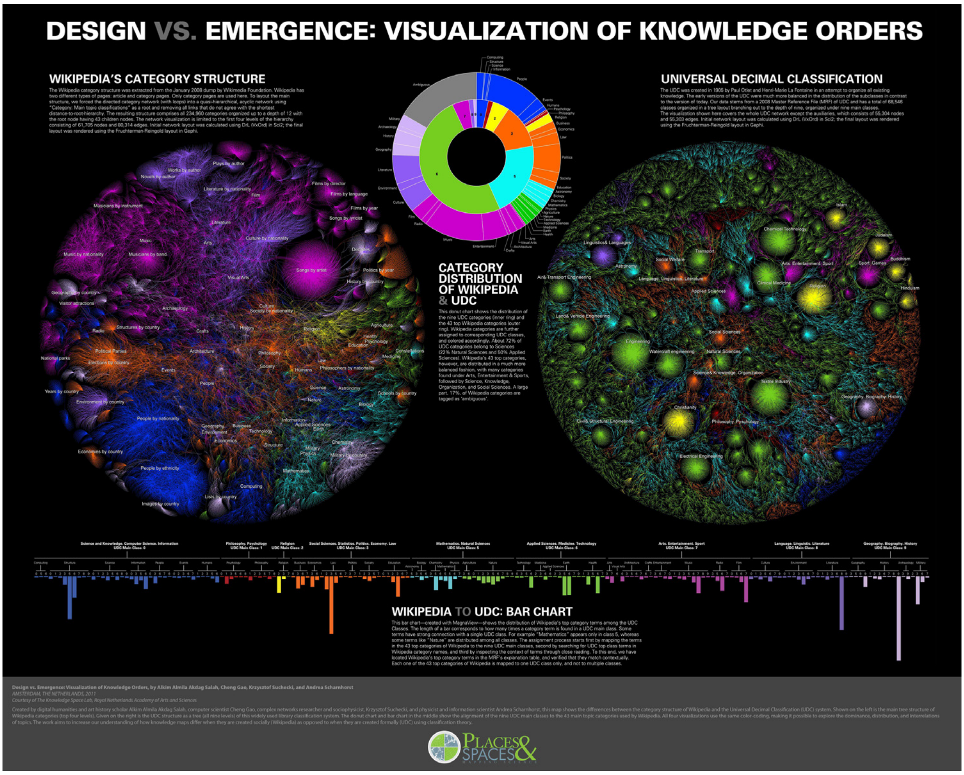


Figure 5. Wikipedia's category structure (left) versus the Universal Decimal Classification (right).[9]
Source. Salah, A. A. A., Gao, C., Scharnhorst, A., & Suchecki, K. (2011). Design vs. emergence: Visualisation of knowledge orders. Courtesy of The Knowledge Space Lab, Royal Netherlands Academy of Arts and Sciences. In K. Börner & M. J. Stamper (Eds.), *7th iteration (2011): Science maps as visual interfaces to digital libraries –Places & spaces: Mapping science*. Available from http://scimaps.org Reproduced in Börner, K. (2015). *Atlas of Knowledge: Anyone can Map*. Cambridge, MA: MIT Press.

processing (Waltz, 2014), and image processing and video analysis (Klette, 2014). These techniques involve mathematics and computational algorithms for natural language understanding and motion tracking, video tracking, shape recognition, object detection and facial recognition in computer science and artificial intelligence to identify people, (un)usual behavior and situational awareness in a wide range of domains, including surveillance, safety and security, and health care, for example. Critical here, as in the research framework as a whole, is the ability to describe the impact of *combinations* of choices from different areas and modalities in different contexts. However, a major challenge remains the provision of contextual information to aid the analytical and interpretative process, to which we now turn.

Using Wikipedia for Contextual Information

The multimodal research framework addresses the issue of complexity by using a social semiotic approach to multimodal discourse analysis in conjunction with computational modeling approaches, aided by contextual information derived from Wikipedia, a ''socially evolved'' knowledge system which has become a major source of knowledge about human sociocultural

Table 2. Sample Three-Level Categorization Tree for Keyword "World Trade Center."

Keyword	World Trade Center
Categories: Level 1	World Trade Center (2001-present)
Categories: Level 2	World Trade Center; Financial District, Manhattan; history of New York City; office buildings in Manhattan; Port Authority of New York and New Jersey; skyscrapers in Manhattan; skyscrapers over 350 meters; World Trade Centers; Landmarks in Manhattan
Categories: Level 3	Buildings and structures in Manhattan; World Trade Centers; Financial District, Manhattan; Port Authority of New York and New Jersey; Lower Manhattan; neighborhoods in Manhattan; central business districts in the United States; Lower Manhattan; neighborhoods in Manhattan; New York City; histories of cities in New York; history of the New York metropolitan area; buildings and structures in Manhattan; office buildings in New York City; intermodal transportation authorities in New York; intermodal transportation authorities in New Jersey; transportation in New York City; port authorities in the United States; United States interstate agencies; Port of New York and New Jersey; buildings and structures in Manhattan; skyscrapers in New York City; skyscrapers by height; commercial buildings; landmarks in New York City; Manhattan

life. Salah, Almila, Cheng, Krzysztof, and Scharnhorst (2012) explore the dominance, distribution, and interrelations of different topics in the category structures of Wikipedia and the Universal Decimal Classification (UDC) by color coding the various topics (cf. Figure 5). The results show clear differences between the two classification systems: that is, a substantial component of the Wikipedia classification system is devoted to the environment, culture, literature, film, radio, music, entertainment, the arts, architecture, people and events (left, colored purple, pink, and blue), unlike the UDC, a ‘‘designed’’ knowledge system dominated by the applied sciences (right, colored green).[8] That is, Wikipedia, as a socially designed knowledge system, has an extensive coverage of human sociocultural life, compared with the formally designed UDC system.

As Figure 5 shows, Wikipedia provides an extensive classification system for deriving structured information about human life, and by default, what people are communicating via textual (written and spoken) and visual resources (e.g., photographs, drawings and cartoons, etc.). As a result, Wikipedia classifications of location, keywords, and visual objects provide a good basis for interpreting discourse patterns contextually, creating a link between the sign itself (i.e., actual words and images), context, and meaning. For instance, and returning to the image in Figure 1, Wikipedia categories provide higher order semantic information for locations: for example, the Wikipedia categories for ‘‘World Trade Center,’’ in this case, ‘‘World Trade Center (2001-present),’’ provide higher order contextual information (e.g., Financial District, Manhattan; History of New York City; Office buildings in Manhattan; etc., see ‘‘Categories: Level 2’’ in Table 2) for interpreting experiential and interpersonal meanings in the multimodal message. These categories are in turn organized into higher order semantic categories in Wikipedia (see Level 3 in Table 2). There are also Wikipedia categories for ‘‘World Trade Center (1973-2001)’’ which contain categories pertaining to the September 2001 attacks (e.g., Buildings and structures destroyed in the September 11 attacks), the prior history of the site (e.g., Twin Towers), other destroyed buildings in New York and the United States (e.g., Destroyed landmarks in New York), and other semantic categories (e.g., Financial District, Manhattan).

Similarly, key linguistic and visual items in the message itself (e.g., ‘‘CO-TF1’’) provide higher order semantic information (e.g., Urban Search and Rescue Task Forces, Government of

Colorado) for interpreting multimodal messages. Wikipedia classification trees also provide semantic information about visual objects identified through image-processing techniques (e.g., buildings, people). As a socially constructed knowledge system, Wikipedia categories cover most aspects of sociocultural life which form the basis of human discourse. While the contributors to such large collaboratively edited reference projects such as Wikipedia are currently limited (e.g., women, ethnic minorities, and languages other than English are underrepresented: see ''systemic bias'' in Wikipedia[10]), our research framework is presented with a view to the future availability of comprehensive knowledge databases which will reflect the full diversity of human life as the Internet continues to rapidly grow and expand.

Multimodal Mixed Methods Research Framework: Design and Methodology

The multimodal mixed methods research framework involves the development and implementation of a multilevel, multimodal contextual model for analyzing, visualizing, and extracting useful information. The research framework is based on the premise that multimodal social semiotic theory offers: (a) a rich theoretical foundation for identifying key systems for experiential, logical, and interpersonal meaning in different media; (b) that these systems can be interpreted using contextual information, data-mining, and machine-learning techniques; and (c) the resulting discourse patterns can be revealed using information visualization techniques.

The research framework involves specific tasks for integrating qualitative and quantitative approaches for this purpose. These tasks involve: (a) implementing *qualitative* methods: that is, identification of key linguistic and visual systems which operate across different media by using digital tools to transform the qualitative analysis into quantitative data; (b) implementing *quantitative* methods: that is, use of machine learning to develop data-mining algorithms for these systems, using contextual information derived from metadata and socially designed knowledge systems, in this case, Wikipedia; and (c) synthesizing these methods as new algorithms in prototype technology, referred to as the *multimodal analysis visual information system* (MMA–VIS), for exploring the results of the analysis. At present, no techniques exist that truly combine multimodal analysis, data mining, and information visualization simultaneously, due to the inherent complexity and the challenges of disciplinary and theoretical integration. The conceptual framework and methods we employ for this are explained in detail below.

Example of the Multimodal Research Framework Design for Crisis Informatics

The multimodal research framework is described below by applying the framework to public reports of extreme events in Australia. The research framework addresses current gaps in crisis informatics in two ways: first, through the development of conceptual models and analytical tools which integrate qualitative and quantitative methods and second, by using these techniques to develop interactive analytical tools to advance the knowledge and capacity for disaster analysts. That is, the research framework establishes the necessary foundation for creating and exploring new techniques and tools for analyzing, visualizing, and mapping the content (i.e., processes, participants, and circumstances) and interpersonal stance and appraisal of text and images in cross-media reports of extreme events (in this case, online news and social media) in affected areas. The ability to automatically derive meaning from public discourses will enable complex information to be interpreted during times of crisis, and it should enhance understanding of how changes in communication practices arising from new media technologies can be harnessed to provide critical information about extreme events from people in disaster zones.

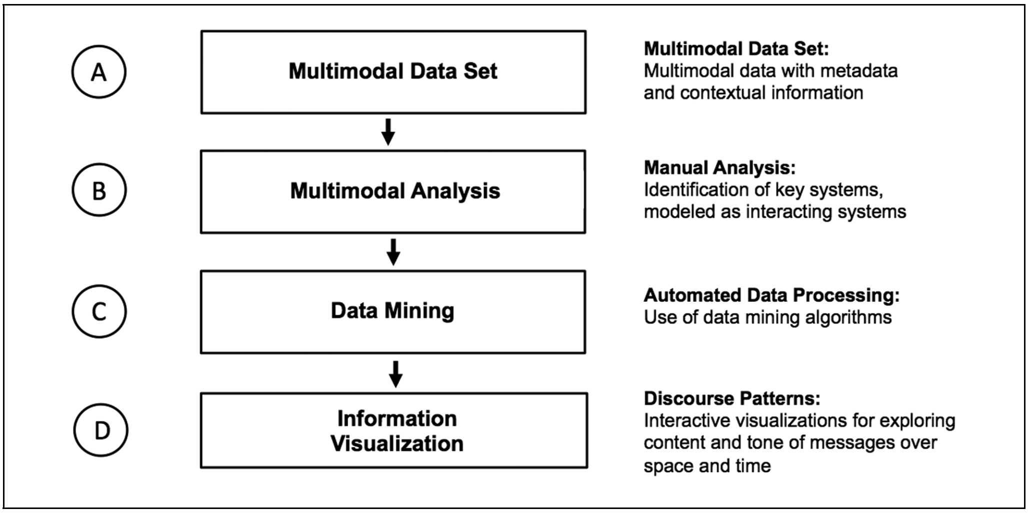

A
Multimodal Data Set
B
Multimodal Analysis
C
Data Mining
D
Information Visualization
Multimodal Data Set: Multimodal data with metadata and contextual information
Manual Analysis: Identification of key systems, modeled as interacting systems
Automated Data Processing: Use of data mining algorithms
Discourse Patterns: Interactive visualizations for exploring content and tone of messages over space and time

Figure 6. Research methodology.

At a more abstract level, this mixed methods approach aims to contribute to our understanding of human behavior when lives are placed at risk.

The multimodal discourse descriptors, associated algorithms for multimodal data mining, and interactive visualizations for modeling and representing discourse patterns are being developed for online news and public discourses in areas affected by extreme events. This approach involves data collection, data contextualization, manual annotation, automated data-mining techniques, and information visualization techniques. The research methodology A to D, displayed in Figure 6, is described below.

A. Multimodal Data Set (Data Collection Phase). The multimodal data set comprises public discourses about extreme events collected from major online Australian news and social media sourced from affected areas. Contextual information derived from metadata (e.g., time, geographical coordinates, user information) and Wikipedia classifications of locations, keywords, and visual objects which have been processed using computer vision methods such as object recognition and categorization, are then incorporated in the multimodal data set. The Wikipedia categorization structures provide reasonable semantic approximations of location and other data, providing contextual information for interpreting the geotagged data sets. This is shown in O'Halloran, Chua, et al.'s (2014) study of the role of images in social media messages in urban Singapore, where Wikipedia categorizations of Foursquare venues provided semantic descriptions of social activities in various locations, permitting multimodal messages to be linked to social context. Although not always precise, Wikipedia categorizations provided reasonably robust semantic approximations, revealing how linguistic and visual selections in social media messages vary across different social contexts. As O'Halloran, Chua, et al. (2014) explain, the methodology involves using DBPedia[11] to organize a structured representation of Wikipedia classification trees (i.e., article categories and categories of category mappings). Using this approach, location information (e.g., Foursquare venues) and key linguistic and visual selections which are associated with a Wikipedia page are mapped onto higher level semantic categories, providing the contextual information for interpreting the geotagged data in the multimodal data set.

B. Multimodal Analysis (Qualitative to Quantitative Data Transformation Phase). Close, qualitative multimodal analysis of selected multimodal texts is undertaken using multimodal annotation software (e.g., see Figure 3). The resulting structured quantitative data for text and images (see example in Figure 3a) consists of the text identification code, the system catalogue identification code, the system name, the set of options in the system network organized according to hierarchical order, the selected system choice, the overlay type for the system choice (i.e., rectangle, circle, polygon, line, or pin), the coordinates of the overlay, and the textual descriptors, as summarized below:

> Text/Image—Text Name, System Catalogue, System Name, System Network, Selected System Choice, Overlay Type, Points X-Y, Description

The structured data for videos (see example in Figure 3b) consist of the text identification code, the system catalogue identification code, the subcategory (for the speaker, visual track, soundtrack, etc.), the name of the system, the set of options in the system network organized according to hierarchical order, the selected system choice, the absolute start time of the system choice, the absolute end time of the system choice, the absolute duration of the system choice, the relative start time of the system choice (in %), the relative duration of the system choice (in %), and the transcription, as summarized below:

> Video—Text Name, Catalogue, Subcategory, System Name, System Network, Selected System Choice, Absolute Start Time, Absolute End Time, Absolute Duration, Relative Start Time (in %), Relative Duration (in %), Transcription

There is a common reference to the original media files in terms of spatial coordinates (for text and images) and time stamps (for video), so the results can be used to inform the subsequent development of data-mining algorithms.

C. Data Mining (Quantitative Phase). The multimodal research framework integrates the manual analysis undertaken using the multimodal analysis software applications and the contextual information derived from the metadata and Wikipedia classifications with data-mining techniques. The approach involves the development of a new semantically focused, multiscale, cross-media model that addresses the current limitations of existing approaches such as bag-of-words modeling (where the semantic relations between elements are not addressed) and its basic probabilistic extensions (Wallach, 2006). That is, by exploiting advances in multimodal analysis, natural language processing, image and video processing and by making use of Wikipedia, the research framework introduces multiple semantic layers over the raw data. In this pyramid-style modeling, textual and visual data are analyzed semantically according to the suitable context to extract high-level semantic terms (e.g., named and visualized entities). Using information available from multimodal analysis and exploiting the hierarchical information provided by Wikipedia, the subsequent higher layers are designed to contain more generic, categorical abstractions of the content of lower layers. Such multiscale modeling (e.g., Tang et al., 2013) guides subsequent data-mining tasks and facilitates visualizations for varying degrees of semantic depth.

D. Information Visualization (Exploratory Phase). The results of the data mining are displayed using the prototype software application, MMA–VIS, which has functionalities to support the representation of discourse patterns to the various degrees of semantic depth derived from the multiscale multimodal data-mining techniques. These varying degrees of depth include, for example, distinctions within the expression plane (i.e., actual word, image, video), among the metadata (e.g., user, time, location), and in topic and content relations (e.g., named and

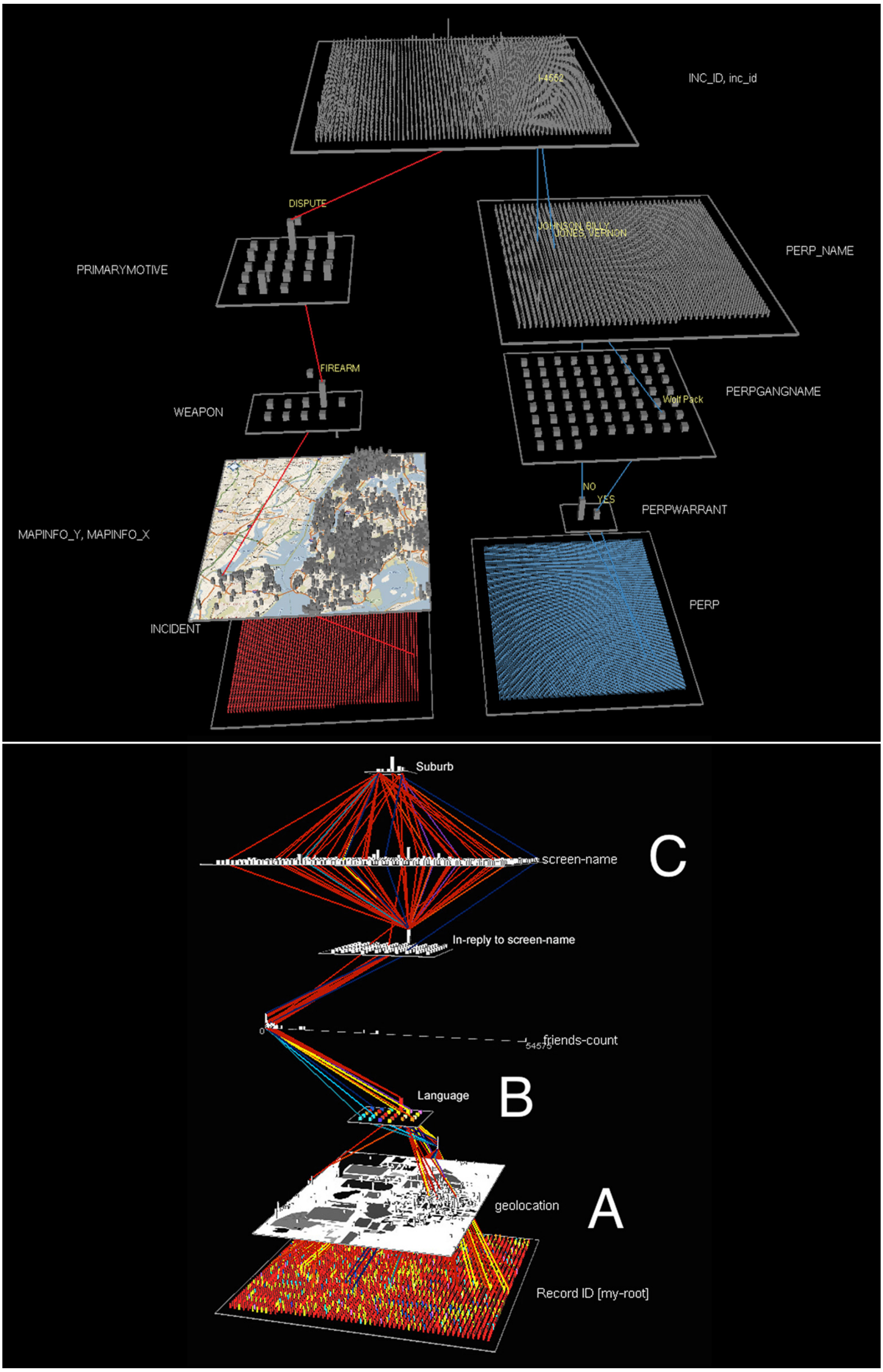


Figure 7. Linked arrays in Starlight Visual Information System.

Note. Top: Image courtesy of the Pacific Northwest National Laboratory, operated by Battelle Memorial Institute for the U.S. Department of Energy. Bottom: Image produced using Starlight, a software product created at the Pacific Northwest National Laboratory.

visualized entities) and interpersonal meaning (i.e., speech function, appraisal, modality, emotion). The point of departure for MMA–VIS is Starlight Visual Information System,[12] which contains platforms for importing disparate data sets, data processing, data integration, and visual analysis and reporting of the results.

Starlight VIS facilities include linked arrays (Figure 7) for identifying relations across multiple dimensions (e.g., message, location, extracted features, and intertextual relations with other data sets; Figure 7; top). For example, in a 2014 pilot project employing the Starlight system, a set of geospatial coordinates was used to define a physical area, in this case Curtin University in Perth Western Australia. Multimodal data were collected from Twitter, Instagram, and selected web pages at the university. In this case, the linked array (Figure 7; bottom) shows the relations between multiple levels in the data model: that is, individual tweets and their location at Curtin University (labeled A), the language (e.g., English, Indonesian, and Thai; labeled B), and the user networks (labeled C). In the multimodal research framework articulated here, linked arrays are used to identify patterns derived from the manual multimodal analyses, Wikipedia semantic descriptions and automated data processing, in order to refine the data-mining techniques and develop new visualization techniques to facilitate the modeling of discourse patterns, based on these results.

Visualization techniques assist both in the determination of features out of often complex multidimensional analyses and the interpretation of large-scale automated recognition analyses (Munzner, 2014). MMA–VIS is intended to provide visual access to the patterns derived from the multimodal data analysis and automated techniques, in ways which will make the relations and patterns evident to the viewer through careful, motivated style, and design choices (e.g., Vande Moere, Tomitsch, Wimmer, Boesch, & Grechenig, 2012). MMA–VIS is being developed to have facilities for exploring data and investigating their intertextual relations (e.g., according to discourse type, semantic level, location, time, and other metadata) across different temporal and physical scales through visual correlations of the match between expected discourse patterns of experiential and interpersonal meanings and resultant discourse patterns, based on the algorithms developed in trial studies. The results are displayed on a map featuring hierarchical layers of semantic analysis that can be manipulated, together with a platform for comparing patterns from individual texts to larger scale patterns.

Theoretical and Analytical Contributions of the Multimodal Research Framework

The multimodal mixed methods research framework presented in this article incorporates theoretical contributions to discourse analysis and technological tools for analyzing and visualizing public discourse. As such, the work complements existing research in cultural and new media studies with a computational social science approach that employs social semiotic theory for multimodal discourse analysis. In particular, the research framework moves beyond current automated techniques (e.g., topic, sentiment, and social network analysis), to visualizing and mapping linguistic and visual content (in terms of happenings and the logical relations between those happenings), and interpersonal meaning (i.e., interpersonal stance to and appraisal of happenings) of the messages. A critical advance is the deployment of results from social semiotic register theory that document how key systems of meaning (linguistic and visual) are activated according to discourse type and context. This provides the potential for major breakthroughs in big data analytics by guiding the construction of reliable correlations between observable data and their social import.

The research framework also contributes to the advancement of knowledge from a data-mining perspective, as current information visualization models are limited to volumes (e.g., tracking of hash-tags in Twitter) or word clouds, which lack the additional layer of semantic information, derived from socially evolved classification systems such as Wikipedia. In this regard, the research framework is an extension of prior work, resulting in hierarchical layers of semantic information that allow connections to be made between the import of words/terms and their contextual relations, providing a basis for next-generation visualization mechanisms. These insights will result in a new capacity for developing multimodal descriptors for visualizing, tracking, and mapping significant patterns of meaning in diverse data sets, adding a social layer to Digital Earth (Jackson & Simpson, 2013).

The multimodal research framework is an example of the digital mixed methods design which extends advanced mixed methods research into the realm of data mining and information visualization for big data analytics. The advanced mixed methods design moves beyond data merging using standard statistical techniques and graphs and text-based displays of information (e.g., joint displays) into the realm of interactive, visual representations of abstract data. Building on theory from information design, computer graphics, human–computer interaction, and cognitive science which recognizes that humans have remarkable perceptual abilities, information visualization permits researchers to explore patterns in large, multidimensional data sets in new ways. By being able to undertake tasks such as overviewing the whole data set, zooming into items of interest, filtering out items, selecting details-on-demand, extracting subcollections, and keeping a history of actions (Shneiderman, 1996) in different visualization types—for example, 1D (linear), 2D (planar, including geospatial visualizations), 3D (volumetric), temporal, *n*-Dimensional, tree/hierarchical, network diagrams and linked arrays—mixed methods researchers are able to explore structures, make connections and establish causal relations in ways which are currently not possible. The digital mixed methods design thus makes a contribution to the interdisciplinary study of big (digital) sociocultural data sets in fields which include the social sciences, education, psychology, and health sciences, as well as the computational sciences.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support received from Curtin University for the 2014 pilot project which used Starlight Visualization Information System software.

Notes

1. https://upload.wikimedia.org/wikipedia/commons/7/70/FEMA_Colorado_Task_Force_at_WTC.jpg
2. http://www.qsrinternational.com/
3. http://atlasti.com/
4. https://tla.mpi.nl/tools/tla-tools/elan/
5. http://multimodal-analysis.com/products/multimodal-analysis-image/software/
6. http://multimodal-analysis.com/products/multimodal-analysis-video/
7. http://www.curtin.edu.au/ (retrieved 5 September 2015)
8. http://www.curtin.edu.au/livingourvalues/about/senior-leaders-conference-15-16-august-2013.cfm
9. http://scimaps.org/mapdetail/design_vs_emergence__127

10. https://en.wikipedia.org/wiki/Wikipedia:Systemic_bias#cite_note-2
11. http://wiki.dbpedia.org/
12. http://starlight.pnnl.gov/

References

Bateman, J. (2014a). Looking for what counts in film analysis: A programme of empirical research. In D. Machin (Ed.), *Visual communication* (pp. 301-329). Berlin, Germany: De Gruyter Mouton.

Bateman, J. (2014b). Using multimodal corpora for empirical research. In C. Jewitt (Ed.), *The Routledge handbook of mutimodal analysis* (pp. 238-252). New York, NY: Routledge.

Bateman, J. (2016). Methodological and theoretical issues for the empirical investigation of multimodality. In N.-M. Klug & H. Stöckl (Eds.), *Sprache im Multimodalen Kontext* [Language and multimodality]. Berlin, Germany: De Gruyter Mouton.

Bazeley, P. (2010). Computer assisted integration of mixed methods data sources and analyses. In A. Tashakkori & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research* (2nd ed., pp. 431-467). Thousand Oaks, CA: Sage.

Cao, N., Lin, Y.-R., Sun, X., Lazer, D., Liu, S., & Qu, H. (2012). Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, *18*, 2649-2658.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. London, England: Sage.

Curry, L., & Nunez-Smith, M. (2015). *Mixed methods in health sciences research: A practical primer* (Vol. *1*). Thousand Oaks, CA: Sage.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research, 48*(6 Pt. 2), 2134-2156.

Gardner, S. (2012a). A pedagogic and professional case study: Genre and register continuum in business and in medicine. *Journal of Applied Linguistics and Professional Practice*, *9*, 13-35.

Gardner, S. (2012b). Genres and registers of student report writing: An SFL perspective on texts and practices. *Journal of English for Academic Purposes*, *11*, 52-63.

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Annals of Family Medicine*, *13*, 554-561.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London, England: Edward Arnold.

Halliday, M. A. K. (2002). Text as semantic choice in social contexts (1977). In J. Webster (Ed.), *Linguistic studies of text and discourse: Volume 2 in the collected works of M. A. K. Halliday* (pp. 23-81). New York, NY: Continuum.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Geelong, Victoria, Australia: Deakin University Press [Republished by Oxford University Press 1989].

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London, England: Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). New York, NY: Routledge.

Hatim, B., & Mason, I. (1990). *Discourse and the translator*. New York, NY: Routledge.

Jackson, D., & Simpson, R. (2013). *D_City: Digital Earth: Virtual nations: Data cities*. Newtown, New South Wales, Australia: DCity Pty.

Jewitt, C. (2014). Different approaches to multimodality. In C. Jewitt (Ed.), *Handbook of multimodal analysis* (pp. 31-43). New York, NY: Routledge.

Klette, R. (2014). *Concise computer vision*. London, England: Springer-Verlag.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London, England: Routledge.

Lukin, A., Moore, A. R., Herke, M., Wegener, R., & Wu, C. (2011). Halliday's model of register revisited and explored. *Linguistics and the Human Sciences*, *4*, 187-213.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., & . . .Blanford, J. (2011, October). Senseplace2: Geotwitter analytics support for situational awareness. *Visual Analytics Science and Technology (VAST) 2011 IEEE Conference* (pp. 181-190). Boulder, CO: IEEE.

Manovich, L. (2012). How to compare one million images? In D. Berry (Ed.), *Understanding digital humanities* (pp. 249-278). London, England: Macmillan.

Margolis, E., & Pauwels, L. (Eds.). (2011). *Sage handbook of visual research methods*. London, England: Sage.

Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed.). London, England: Continuum.

Matthiessen, C. M. I. M. (2009). Multisemiotic and context-based register typology: Registerial variation in the complementarity of semiotic systems. In E. Ventola & A. J. Moya (Eds.), *The world told and the world shown: Multisemiotic issues* (pp. 11-38). Hampshire, England: Palgrave Macmillan.

Matthiessen, C. M. I. M. (2013). Applying systemic functional linguistics in healthcare contexts. *Text & Talk*, *33*, 437-467.

Morse, J., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.

Munzner, T. (2014). *Visualization analysis & design*. New York, NY: CRC Press.

O'Halloran, K. L., Chua, A., & Podlasov, A. (2014). The role of images in social media analytics: A multimodal digital humanities approach. In D. Machin (Ed.), *Visual communication* (pp. 565-588). Berlin, Germany: De Gruyter Mouton.

O'Halloran, K. L., E, M. K. L., Podlasov, A., & Tan, S. (2013). Multimodal digital semiotics: The interaction of language with other resources. *Text & Talk*, *33*, 665-690.

O'Halloran, K. L., E, M. K. L., & Tan, S. (2014). Multimodal analytics: Software and visualization techniques for analyzing and interpreting multimodal data. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (2nd ed., pp. 386-396). London, England: Routledge.

O'Halloran, K. L., E, M. K. L., & Tan, S. (2015). Multimodal semiosis and semiotics. In J. Webster (Ed.), *The Bloomsbury companion to M. A. K. Halliday* (pp. 386-411). New York, NY: Bloomsbury.

O'Halloran, K. L., Wignell, P., & Tan, S. (2015). Multimodal social semiotic approaches to analyzing and designing brand communications. In G. Rossolatos (Ed.), *Handbook of brand semiotics* (pp. 280-328). Kassel, Germany: Kassel University Press.

O'Toole, M. (2011). *The language of displayed art* (2nd ed.). New York, NY: Routledge.

Pipek, V., Liu, S. B., & Kerne, A. (2014). Crisis informatics and collaboration: A brief introduction. *Computer Supported Cooperative Work (CSCW)*, *23*, 339-345.

Salah, A., Almila, A., Cheng, G., Krzysztof, S., & Scharnhorst, A. (2012). Need to categorize: A comparative look at the categories of Universal Decimal Classification System and Wikipedia. *Leonardo*, *45*(1), 84-85.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of IEEE Symposium on Visual Languages* (pp. 336-343). Boulder, CO: IEEE.

Steiner, E., & Yallop, C. (Eds.). (2001). *Exploring translation and multilingual text production: Beyond content*. Berlin, Germany: Walter de Gruyter.

Tan, S., Smith, B. A., & O'Halloran, K. L. (2015). Online leadership discourse in higher education: A digital multimodal discourse perspective. *Discourse & Communication*, *9*, 559-584.

Tang, H., Shen, L., Qi, Y., Chen, Y., Shu, Y., Li, J., & Clausi, D. A. (2013). A multiscale latent dirichlet allocation model for object-oriented clustering of VHR panchromatic satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *51*, 1680-1692.

van Leeuwen, T. (1999). *Speech, music, sound*. London, England: Macmillan.

Vande Moere, A., Tomitsch, M., Wimmer, C., Boesch, C., & Grechenig, T. (2012). Evaluating the effect of style in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *18*, 2739-2748.

Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 977-984). New York, NY: ACM.

Waltz, D. L. (Ed.). (2014). *Semantic structures: Advances in natural language processing* (Vol. *23*). New York, NY: Routledge.