Suddenly, multimodality is a hot academic topic. University departments that for decades studied the honourable fields of language & literature – and if adventurous, their interrelations – now rapidly begin to change tack. Whether this is primarily because those in charge realize that the art and communication of the near future will be ever less exclusively verbal in nature or, more banal, because they see their institutional resources dwindle along with their student enrolment – that question shall here remain unanswered.

Anthony Baldry and Paul Thibault present a course book on the transcription and analysis of multimodal text, rooted in Hallidayan semiotics, which connects this monograph to the work of Gunther Kress and Theo van Leeuwen who, also in 2006, launched a modestly revised version of their daring, but not uncontroversial *The Grammar of Visual Design* (Kress and Van Leeuwen, 1996/2006; for critical discussion, see Forceville, 1999; Bateman et al., 2004). Baldry and Thibault shy away from defining the thorny concept of “multimodality,” taking as their starting point that “there are many other resources that can be used to create texts in addition to the spoken and written word” (p. 4). It transpires that they take the various “semiotic systems” that make up multimodal texts to comprise at least language, gesture, music, and movement (p.1; for another attempt at definition see Kress and Van Leeuwen, 2001). And although “semiotics” to some humanities scholars may have the ring of yesteryear’s scholarly fashion, there is no doubt that semioticians have always shown a theoretical interest in multimodal texts, so that they have an advantage now multimodality is acknowledged as a subject highly pertinent to representation in the 21st century.

Baldry and Thibault’s book consists of an introductory chapter on the importance of genre for the analysis of multimodal discourse, followed by three chapters on printed, web, and filmic texts, respectively, focusing in great detail on some twenty case studies. The book contains 23 tables, 3 full-blown transcripts of TV commercials, almost 50 figures, and 19 boxed “Inserts” that focus on key concepts (e.g., “metafunctions,” “clusters and cluster analysis,” “projection”) and that are scattered throughout the book. Finally, a website is presented, with the promise of further exercises.

The central ideas of the book are that the various modes operative in a multimodal text need always be considered in the way they interact (an idea Baldry and Thibault label the “resource integration principle,” p. 4), and the “meaning-compression principle,” which “refers to the effect of the interaction of smaller-scale semiotic resources on higher-scalar levels where meaning is observed and interpreted” (p. 19). Following the Hallidayan paradigm, the authors elaborate on the experiential, interpersonal, textual, and logical dimensions of language, intending to demonstrate that these work for other sign systems as well (p. 22).
One problem besetting the theorization of multimodal discourse is that most senior scholars entering this field have been monomodally educated: they are linguists, or musicologists, or art historians. Inevitably, they are thereby biased by their original field of study, and limited by their restricted knowledge of other disciplines. But inasmuch as language is the most detailed and refined mode for the communication of complex information, linguistics is not at all a bad discipline to start working from. Baldry and Thibault’s basic approach is to delimitate the variables in different semiotic modes, or sign systems, that will or may affect the interpretation of the multimodal text in which they appear. With painstaking precision they describe their case study texts, listing a long array of verbal, visual, musical, gestural, and sonic details, using numerous italicised technical terms, and constantly referring the reader to the various Inserts. Sometimes this yields interesting local insights in a text’s design and presumed impact. However, these descriptions seldom result in non-trivial explanations why the texts convey what they supposedly do convey, let alone in the formulation of – however tentative – patterns or generalizations. By and by these long descriptions become a real chore to read, and often do not exceed the level of truisms (“the mouse can be used to perform the following kinds of actions on objects: (1) point to object; (2) roll over object; (3) click object,” p. 153).

Bluntly speaking, the authorial descriptions procure insufficient insights that a moderately attentive viewer-reader of the case-study under discussion had not already grasped himself, and that does not make for exciting reading. The disappointment can be at least partially ascribed to the fact that Baldry and Thibault’s orientation is too much bottom-up. I am a great advocate of textual analyses, but these must be complemented by top-down conceptualisations to avoid infinite detail, and unfortunately neither the four Hallidayan dimensions, nor the numerous “key concepts” in the Inserts, are shown to be able to perform the hoped-for synthesis. The need for well-functioning top-down conceptualisations – arguably always needed in scholarship – is especially sharply felt in the theorizing of multimodal texts. Since the producer of a multimodal text has two or more different modes at her disposal to convey a coherent set of snippets of facts and emotions, she often can choose between modes for each of the snippets. In a film sequence, for instance, a director can command a character to say: “I am angry”; make him clench his teeth and frown; and order him to start howling angrily – to mention only three modes (language, bodily expression, sound) that can be used both separately or in combination. But for this to be theorizable, it is necessary to refer to supra-textual, internalised models that humans have at their disposal. In this respect, there are a number of approaches that I have found helpful, and that Baldry & Thibault either do not mention or do not discuss in much detail, presumably because they play no role in the Hallidayan approach. These include genre – which is mentioned by the authors, but not really made productive in the analyses – narrativity, Relevance Theory, the notion of Idealized Cognitive Models and prototype theory (Lakoff, 1987), and Fauconnier & Turner’s (2002) Blending Theory (which shares elements with Baldry and Thibault’s resource integration and meaning-compression principles; but cf Forceville, 2004). For more discussion and references, see Forceville (1996, 2006).

A second reason for the limited appeal of the book, alluded to above, is that Baldry and Thibault’s expertise is rather heavily language- and social semiotics-oriented, as transpires from their bibliography. Although they admirably produce variables pertinent to the non-language modes, it is a pity they are completely oblivious to fine work done by experts in other disciplines. Why re-invent the wheel for the analysis of cinematic footage when there is Bordwell and Thompson’s excellent and wonderfully written Film Art (continually revised since its first edition in 1979, the eighth edition launched in 2006)? Drawing on a textbook such as Bordwell and Thompson’s would have benefited Baldry and Thibault in the use of accepted film terminology as well as in describing the phenomena to which these terms refer. To give a few examples of standard concepts in the analysis of visual
footage and sound in film that Baldry and Thibault either do not discuss at all or describe very laboriously in their own terms: the 180 degrees rule, Hollywood continuity editing, shot-reverse shot, cross-cutting, cut-ins, score versus source music, off-screen versus voice-over text, temporal ellipsis, mickey-mousing, eyeline matches … Similarly, the authors are not cognizant of work done on comics. If they had studied McCloud (1993; see also McCloud, 2006), they might have realized that their analysis of the Lupo Alberto panels (figure 1.7, p. 36) lacks a discussion of the meaning of the “gutter” – the space that separates two panels – and of the non-real hearts surrounding the hen’s head. In fact, these hearts are not mentioned at all by the authors, while they probably contribute more to our awareness that the hen is in love with Alberto than either the eye position or the posture emphasized by Baldry and Thibault. In a section advising on the systematic study of tables, there is no mention, let alone discussion, of insights by leading scholars in the field of graphic design and visual maps such as Jacques Bertin and Edward Tufte (references to their work as well as others can be found in Engelhardt’s (2002) synthesizing study on the multimodal semantics and syntax of maps, charts, and diagrams). And a pioneering study such as Manovich (2001), an author who himself combines expertise in New Media with knowledge of film, fiction, and art, might have provided the kind of concepts to structure and put into perspective Baldry and Thibault’s analyses of web pages (see for instance the difference between database logic and narrative logic, Manovich 2001: 218ff, and the section on “ navigable space,” Ibid. p. 244ff.).

Clearly, nobody can read everything, and multimodality is, as Baldry and Thibault acknowledge, a discipline that is as young as it is vast. But surely one could have reasonably expected them to have familiarized themselves first-hand with at least some important work done by experts on other modes and media than the ones they themselves are specialists in – even if, or precisely because of the fact that, these experts do not work in a Hallidayan framework. For the reasons outlined above, I therefore believe that Baldry and Thibault’s is not the best book to despatch students onto the vast ocean of multimodal discourse. That best book, to my knowledge, does not yet exist, and is not likely to be written for quite a while. I think the ambitions for the near future should be more modest. It is crucial that scholars embarking on the field of multimodality possess or acquire more than passing knowledge of at least one mode outside the one they have been primarily trained in. With that stereo-knowledge as venture capital, they should undertake studies that investigate series of texts that are alike in important respects besides the, say, two or three modes these texts deploy: for instance, in the genre (e.g., Altman, 1999; Caballero, 2006; Visch 2007; El Refaie, forthcoming) to which they belong; in the medium they make use of; in the culture in which they flourish; in the multimodal trope they communicate by – and preferably in all of them at the same time. As a starting point, multimodal text corpora need to be defined narrowly in order to enable the identification of significant clusters of variables within and across modes (for instance: “print advertising in 1990s glossy magazines aimed at upper-class women”; “Splasher horror movies of the 1990s”; “Websites promoting art museums”). Only by reducing the number of variables to be taken into account and subsequently analysing their occurrence in precisely circumscribed corpora may the long and laborious work of multimodal text description that Baldry and Thibault promote lead to the discovery which clusters form significant patterns in multimodal texts, and how. And once significant patterns are detected, comparative studies become feasible, while humanities scholars so inclined can start contacting their social science colleagues to jointly embark on empirical research.

References

Bateman, John, Delin, Judy, Henschel, Renate, 2004. Multimodality and empiricism: preparing for a corpus-based approach to the study of multimodal meaning-making. In:


Visch, Valentijn, 2007. Looking for Genres: The Effect of Film Figure Movement on Genre Recognition. PhD thesis, Faculty of Arts, Vrije Universiteit Amsterdam.