DOCUMENT DESIGN COMPANION SERIES

Perspectives on Multimodality

Edited by
Eija Ventola,
Cassily Charles and
Martin Kaltenbacher

Perspectives on Multimodality

Document Design Companion Series

DOCUMENT DESIGN COMPANION SERIES accompanies the journal DOCUMENT DESIGN and focuses on internal and external communication of medium sized to multinational corporations, governmental bodies, non-profit organizations, as well as media, health care, educational and legal institutions, etc.

The series promotes works that combine aspects of (electronic) discourse — written, spoken *and* visual — with aspects of text quality (function, institutional setting, culture). They are problem driven, methodologically innovative, and focused on effectivity of communication. All manuscripts are peer reviewed.

DOCUMENT DESIGN is 'designed' for: information managers, researchers in discourse studies and organization studies, text analysts, and communication specialists.

Editorial Address

Jan Renkema
Tilburg University
Discourse Studies Group
P.O.Box 90153
NL 5000 LE TILBURG
The Netherlands

E-mail: J.Renkema@uvt.nl

Volume 6

Perspectives on Multimodality
Edited by Eija Ventola, Cassily Charles and Martin Kaltenbacher

Perspectives on Multimodality

Edited by

Eija Ventola

Cassily Charles

Martin Kaltenbacher

University of Salzburg

John Benjamins Publishing Company

Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Perspectives on Multimodality / edited by Eija Ventola, Cassily Charles and Martin Kaltenbacher.
p. cm. (Document Design Companion Series, ISSN 1568–1963 ; v. 6)
Includes bibliographical references and indexes.
1. Semiotics. 2. Modality (Linguistics) I. Ventola, Eija. II. Charles, Cassily. III. Kaltenbacher, Martin, 1966- IV. Series.

P99.4.M6P47 2004
401'.41--dc22 2004046238
ISBN 90 272 3206 7 (Eur.) / 1 58811 595 X (US) (Hb; alk. paper)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Contributors

John Bateman
Fachbereich 10
Sprach- und Literaturwissenschaften
Universität Bremen
Bibliothekstr. GW2
D-28334 Bremen
Germany
bateman@uni-bremen.de

Kristin Bührig
Lutterothstr. 5
20255 Hamburg
Germany
Buehrig@pragmatiknetz.de

Cassily Charles
Unit 5/2 Ben Boyd Rd.
Neutral Bay
NSW 2089 Australia
ccha7549@bigpond.net.au

Judy Delin
School of Modern Languages
and Cultures
University of Leeds
LS2 9JT Leeds
United Kingdom
j.delin@leeds.ac.uk

Eva Martha Eckkrammer
Universität Salzburg
Fachbereich Romanistik
Akademiestr. 24
5020 Salzburg
Austria
eva.eckkrammer@sbg.ac.at

Renate Henschel
Fachbereich 10
Sprach- und Literaturwissenschaften
Universität Bremen
Bibliothekstr. GW 2
D-28334 Bremen
Germany
rhenschel@uni-bremen.de

Andrea Hofinger
Universität Salzburg
Fachbereich Anglistik
Akademiestr. 24
5020 Salzburg
Austria
andrea_hofinger@hotmail.com

Klaus Kaindl
University of Vienna
Zentrum fuer Translationswissenschaft
Gymnasiumstr. 50
A 1190 Wien
Austria
klaus.kaindl@univie.ac.at

Martin Kaltenbacher
Universität Salzburg
Fachbereich Anglistik
Akademiestr. 24
5020 Salzburg
Austria
Martin Kaltenbacher@sbg.ac.at

Victor Lim Fei
Blk 339 Jurong East Ave 1 #02-1534
Singapore 600339
volflain@hotmail.com

Peter Muntigl
Universität Salzburg
Fachbereich Anglistik
Akademiestr. 24
5020 Salzburg
Austria
Peter.Muntigl@sbg.ac.at

Kay O'Halloran
Department of English Language
and Literature
National University of Singapore
Block AS5, 7 Arts Link
Singapore 117570
ellkoh@nus.edu.sg

Markus Rheindorf
Stolberggasse 21/32
1050 Wien
Austria
markusr@chello.at

Hartmut Stöckl
Steinstraße 3
04275 Leipzig
Germany
hartmut.stoeckl@phil.tu-chemnitz.de

Christopher Taylor
Centro Linguistico d'Ateneo
via Filzi 14
34123 Trieste
Italy
taylor@sslmit.univ.trieste.it

Eija Ventola
University of Helsinki
Department of English
P.O. Box 24 (Unioninkatu 40)
FIN-00014 Helsinki
Finland
eija.ventola@helsinki.fi

Acknowledgements

We wish to thank Professor Heinrich Schmidinger, Rector of the University of Salzburg, for providing the funds to invite distinguished international scholars to Salzburg to discuss new perspectives on multimodality in 2002. The present book is a direct outcome of these fruitful discussions. Furthermore, we would like to thank the *Stiftungs- und Förderungsgesellschaft der Paris-Lodron Universität Salzburg* for providing the necessary financial means for the preparation of this publication. Finally, we thank Dr. Anja Schwarzl for formatting the manuscript.

Introduction

Eija Ventola, Cassily Charles and Martin Kaltenbacher
University of Salzburg, Austria

For a growing number of researchers of text and discourse it has over the last decades become increasingly evident, that in the pursuit of understanding communication patterns around us, the analysis of language alone is not enough. The media of communication or, more accurately, the media which provide analysts with an object of study have for the last several millennia permitted a more or less comfortable excision of language, as text/discourse, from its context. Similarly, it has been relatively easy to analyse language and its instantiation in discourse separately from other forms of meaning-making, such as the gestures and facial expressions of a conversation, the illustrations and print of a document, the music and lighting of a drama. The first displacement, that of text from context, has been addressed, or redressed, by both applied and theoretical linguists in many traditions and for many purposes, particularly in the last half-century. This has been done either through building social context into the fundamental structure of linguistic theory (e.g. social semiotic theories of language), or through including contextual observation as a supplement to 'language-only' models. However, the second displacement, that of language from other kinds of meaning-making, has only recently begun to be addressed by linguists.

Although multimodality and multimediality, when seen as combinations of writing, speaking, visualisation, sounds, music, etc., have always been omnipresent in most of the communicative contexts in which humans engage, they have for a long time been ignored, as various academic disciplines have pursued their own research agendas as research fields. Thus, it is relatively recent that the developments of the various possibilities of combining communication modes in the 'new' media, like the computer and the Internet, have forced scholars to think about the particular characteristics of these modes and the way they semiotically function and combine in the modern discourse worlds.

Changes in the technology of communication are forcing the issue for those who would understand the mediation of our lives by discourse. Multimodality, the

interdependence of semiotic resources in text, is no longer so easy to overlook in today's media – particularly in written/visual communication. Indeed, the kind of text where the 'language' can be excised as an independent unit is, in many areas, fast becoming the exception rather than the norm. Accordingly, applications are putting increasing pressure on the development of multimodal theory. The multimodal literacy needs of students are stretching the ability of teachers to support them; information is proliferating in forms which push our methods of sharing it effectively; the shape of discourse communities is changing with the changing shape of texts; and ideological currents are flowing beyond existing linguistic means of analysis and critique.

Linguistic theories of text and discourse have in many respects by now been developed to such a degree of intricacy that the prospect of 'doing it over again' to account equally well for non-linguistic meaning-making is potentially rather daunting. Indeed, some multimodality research takes the approach that what works for language is likely to work for other kinds of meaning-making, and new models need only be adaptations of those already developed for language. Other research looks to expertise in fields outside linguistics – either to disciplines such as musicology or typography, which provide models of specific meaning-making resources other than language, or to such as performance studies and film theory, where language has always been treated together with other kinds of semiosis. Research into multimodality is therefore marked at this point by a broad degree of eclecticism.

This volume therefore represents one signpost on the several paths of multimodality research and theory-building today. The chapters collected here represent a cross-section of current perspectives on multimodal discourse. The book is organised into two parts. Part I focuses on theoretical and methodological issues. Part II covers a wide range of applications for multimodal description.

Part I – Multimodal Issues – starts with a chapter by Hartmut Stöckl, who sets out to develop a system network for modalities by exploring the semiotics of language and image in printed media. He posits all semiotic resources as consisting of core modes (e.g. language), medial variants (e.g. writing), peripheral modes (e.g. typography), sub-modes (e.g. colour) and features (e.g. saturation) and demonstrates that modes can be distinguished through their semiotic properties, their cognitive orientation and their semantic potential. Semiotically, writing operates via two-dimensional arbitrary graphic forms, whereas images mirror concrete three-dimensional objects of the real world. On the cognitive level language successively integrates signs into phrases, while images must be perceived simultaneously and holistically. Semantically language is based on convention and is therefore less polysemous than images, which can only be interpreted in combination with other modes or in a narrowly defined context.

In Chapter 2 Peter Muntigl argues that language and non-linguistic semiotic systems, like gesture, share a similar metafunctional and stratal organisation and are therefore equally grammatically structured. He maintains that gestures can be broken down into a content and an expression plane sharing categories such as Participant, Process or Manner. In addition to that, he claims that the mode of an interaction determines which semiotic systems are activated in a specific situational context, since the mode also determines the channel, medium and language role in a social interaction. Genres that are likely contexts for gestures are those that embody activity sequences and specific participants (stories, procedures, explanations, descriptions) as well as meanings of location or manner.

In Chapter 3 Victor Lim Fei examines the nature of signs within multimodality. Similarly to Muntigl, he argues that visual images operate on a content and an expression plane. Lim claims that the difference in the semiotics of language and visual images lies primarily in the degree of arbitrariness in the relation between signifier and signified. While language signs are arbitrarily chosen to represent their signifieds, visual images are based on a high degree of iconicity. He suggests that icons are the building blocks of visual images which are pieced together to bring across a coherent message. Like language expressions, icons are context and culture dependent, and the objects and ideas they represent may change according to the contextual and cultural background.

In the last chapter of Part I, Chapter 4, John Bateman, Judy Delin and Renate Henschel argue that many established claims about aspects of multimodal discourse are only informal and interpretative, due to the lack of supportive empirical evidence. As a solution they propose multimodal corpus analysis through tools that are specially designed for analytic research into multimodal meaning. They describe first attempts at designing a multimodal annotated corpus, called the GeM (Genre and Multimodality) Model. In this project documents are not only annotated with the well-established tags of traditional corpus linguistics but with new annotation procedures that provide information about the multimodal design and hierarchical structures of individual documents, such as the layout structure of sentences, headings, photos, drawings, figures, captions, text in pictures, icons, table cells, list headers, page numbers, footnotes, etc. Such information can function as an empirical basis for the investigation of the relationship between different document genres and their realizational formats.

Part II – Analyses and Applications – is dedicated to innovative multimodal analyses in a variety of contexts and disciplines. In Chapter 5 Kay O'Halloran investigates how semiotic change in the formal language of mathematics has helped this science to become so effective in describing the physical world. She demonstrates that mathematics originally developed as a multimodal discourse, where the visual played an essential role in establishing meaning in the process of solv-

ing a problem. During the early Renaissance mathematical problems were often visualised through drawn pictures and diagrams to demonstrate the problem but also to engage the students. In the 16th century abstract semiotic metaphors, like lines visualising the path of flying objects, established a significant turning point in mathematics. Symbols representing physical entities became more frequent particularly with the work of Descartes and Newton. Descartes used algebraic formulae for decontextualised representations of curves, and Newton defined scientific phenomena completely through the symbolism of algebraic equations. Modern mathematics is based on complex inter-semiotic relations between language, visual image (graphs, diagrams) and symbolism (algebraic formulae).

In Chapter 6 Martin Kaltenbacher looks into the semiotics of English language teaching CD-Roms and explores which demands text-image combinations have to fulfil in order to provide positive effects for the learners. He claims that many products combine different semiotic modes in a way that may inhibit rather than foster understanding and learning. He analyses the integration of sound waves in so-called pronunciation labs, the semantics of text-picture combinations, and the use of short video clips demonstrating the articulation of model speakers. All visualisations must meet some minimal semiotic requirements: they have to be easily interpretable, they must help the learners disambiguate the meanings to be learnt, and they have to be exact representations of the structures taught. These requirements are often not met, as many visualisations are too general or too ambiguous. As a solution to this, Kaltenbacher proposes replacing complex visualisations with more discrete ones, such as icons.

In Chapter 7 Markus Rheindorf investigates the relations of specific and non-specific modes in the film *Dirty Dancing*. He explores the questions whether the distribution of the modes is significant to the genre dance film, how the modes combine to realise generic structure, and how the genre can be topologically related to other genres. Rheindorf argues that the protagonists of the film construe their class as well as their gender specific identities through the mode dance. Other modes, like dress code and music, support the semiotic functions of dance. Analysing a number of crucial scenes, Rheindorf suggests that set filmic phrases share common patterns and distributions of modes, which keep occurring in certain typical contexts with similar content. Additionally, he argues that the salience of certain filmic scenes is enhanced through a strategic change in their multimodal configurations.

In Chapter 8 Christopher Taylor looks at the role multimodal text analysis can play in the process of subtitling a film in a foreign language. A particular difficulty in this process arises when the spoken text contains a word play which is specific to the source language and is made visible in the film. Here the meaning encoded in the different modes may be lost in the translation process, unless the

translator adopts what Taylor calls "intersemiotic translation". Frequent strategies in subtitling are condensing, decimating, or deleting those elements of the original text which are retrievable to the viewer through the other semiotic modes, like pictures, music or sound. Taylor demonstrates how multimodal film transcription systematically depicts the different semiotic modalities a filmic scene consists of and shows that it can serve as a useful basis for the selection of the verbal elements that should appear in the subtitles.

In the next chapter, Chapter 9, Klaus Kaindl analyses multimodality in the translation of humour in comics, with an emphasis on the comic series Astérix and Tintin. He argues that most of the humour of comics is encoded in the pictures and not in the text. Kaindl identifies eight different strategies for the translation of humour, including deleting humour, changing the modality of humour and inventing absent humour in the translation process. In addition to that, he investigates the relation between verbal text and images in the creation of word and sign play. What creates a particular problem for translation is when a sign play depends on the presentation of an object that belongs to only one culture, e.g., a rugby ball. Furthermore, Kaindl addresses the phenomena of pictorial intertextuality and pictorial allusion and investigates what other modes, like typography, colour or onomatopoeia, add to the humour of comics and how they should be tackled in the translation process.

In Chapter 10 Andrea Hofinger and Eija Ventola analyse the semiotic processes involved in designing a museum on the life of Wolfgang Amadeus Mozart in Salzburg. Hofinger and Ventola study a room in the museum, concentrating on a metafunctional analysis of a painting of Mozart's family and the accompanying audio-text and the display of some furniture in the vicinity of the painting. They focus primarily on how the choices of different ideational, interpersonal and textual resources for the text interrelate to the semiotics foregrounded in the portrait and on what these meanings tell us about the socio-cultural context in which the museum is set up. They demonstrate how careful, or careless, selection of various meaning resources can create interesting and complementary, but sometimes even contrary effects for the overall semiotic of an exhibit.

In Chapter 11 Eva Martha Eckrammer develops a range of theoretical considerations on what she calls the "inter-semiotic layering" of multimodal texts, with a particular focus on medical self-counselling texts and hypertexts. She argues that the pictorial and the verbal elements of a text never develop their meaning separately but through a process of interaction between the semiotic layers involved, though one layer may bear more prominence in the meaning-making than the others. She suggests the following four dimensions of inter-semiotic layering: transposition (image is turned into verbal text or vice versa), juxtaposition (visual and verbal elements coexist without blending), combination (visual and verbal

elements are combined in a text), and fusion (image and verbal text fuse into a new textual form). In addition to that, Eckrammer argues that hypertexts should foster multimodality due to their non-linear structure, although in reality they are conservatively verbal.

In the final chapter, Chapter 12, Kristin Bührig investigates how visual materials, such as diagrams and charts, are integrated in the interpretation process at hospitals, where non-native speakers are briefed about their pending operations by German doctors with the help of a non-professional interpreter. One major obstacle in this type of communication is the linguistic barrier established through the use versus the lack of specialist language and expert knowledge by the parties involved. Bührig focuses on the different discursive roles a diagram plays in the doctor's attempt to communicate knowledge about the operation and the interpreter's attempt to pass on this knowledge in the target language. She finds that the professional doctor uses the visuals systematically to build up the patient's medical knowledge about his diagnosis and for refocusing the thematic elements in his discourse. The interpreter faces problems in transferring the medical information into the target language and uses the diagram to support the new information in the rhematic parts of the translation.

The impetus for this book were discussions which took place during the First International Symposium on Multimodal Discourse at the University of Salzburg in 2002, organised by the editors. We hope that this volume will be a means of opening up the discussion to others who will bring their interest and contribution to the rich and growing area of multimodal discourse.

Salzburg, 1st May, 2004
The editors: Eija Ventola, Cassily Charles and Martin Kaltenbacher

Part I

Multimodal issues

Chapter 1

In between modes

Language and image in printed media

Hartmut Stöckl
Technical University Chemnitz, Germany

The present chapter addresses theoretical and text-analytical issues of the language-image-link in printed media. After modelling multimodality as a networked system of core modes, medial variants, peripheral modes, sub-modes and features, the chapter goes on to characterise the linguistic and the pictorial mode according to semiotic, cognitive and semantic criteria. Central to this chapter is the idea that modes and sub-modes shift or blend (mode overlapping) and mix (mode mixing) in multimodal communicative events. This is demonstrated on two advertising texts sketching out the levels and criteria of analysis necessary to adequately describe language-image-links. The two sample texts show that there is a strong pictorial element in language and a linguistic element in images. The chapter concludes by looking at possible semiotic principles operating across modes.

1. Multimodality – the late discovery of the obvious

Whether as the reflection of a changing communicative landscape (i.e. stronger reliance on modes other than language) or a practical tool in text and discourse analysis, multimodality is currently gaining academic ground. Beyond its mere hype-aspects[1] fostered by an incessant quest for novel research topics, it addresses a phenomenon which is as old as representation itself and crucial to an understanding of almost all forms of communication: multimodal refers to communicative artefacts and processes which combine various sign systems (modes) and whose production and reception calls upon the communicators to semantically and formally interrelate all sign repertoires present. Spontaneous face-to-face talk relies heavily on non-verbal means (e.g. gestures, posture, body language) and has our visually perceptible environment as a constant topic of communication. Often it even has the immediate manipulation of objects as its target. Written language, on the other hand, incorporates images and – through typography and layout – wields strong pictorial powers. More recent media, like film, television or computer, mul-

tiply semiotic potentials by integrating moving images, language (spoken and written), sound and music. The same can already be said of older mediated experience like theatre, drama or opera. I would go as far as to argue that the purely mono-modal text has always been an exception while the core practice in communication has essentially been multimodal all along.[2] The dominance of linguistics, however, and the concentration on language as the central mode, paired with a lack of adequate models for the analysis of other modes, made verbal mono-modality appear to be the standard and dominant form of communication.

Considering that multimodality research has been around only for some 30 years,[3] its achievements are quite remarkable. For one thing, research into multimodality has resulted in various 'grammars' of individual non-linguistic modes.[4] For another, the social semiotic school as one major driving force behind multimodality research has succeeded in showing that 'phylogenetically' some modes seem to displace others in the development of media and genres. Social semiotics has also hinted at possible causes for this,[5] and it has amply illustrated that the specific usage of one or the other mode is guided by socially determined intentions and realises group interests, subjective points of view or ideological stances. A third strand in multimodal research, namely studying the interrelations between various modes, is – as far as I can see – underrepresented. We seem to know more about the functioning of individual modes than about how they interact and are organised in text and discourse.[6]

In the present sketch I want, therefore, to turn towards the interface between two prominent modes, language and image in printed media. First of all, I would like to raise the question of what a mode actually is and how many there are. This involves the tricky business of neatly delineating various sign systems. I hope to show that there is a whole network of heavily interdependent modes and submodes and that in textual practice modes can shift and blend into one another (Sections 2 and 4). Secondly, I seek to demonstrate how language and image can be distinguished from one another looking at their semiotic structure, their semantic characteristics and the cognitive operations they typically entail (Section 3). Thirdly, analysing two sample texts from advertising and describing various facets of inter-modal relations, I will show that there is a pictorial element in language and a linguistic element in images (Sections 4 and 5). Finally, I would like to address the question of whether there are common semiotic principles operating across modes – as Kress and van Leeuwen (2001) have suggested – and what exactly they could be (Section 6).

2. How many modes is 'multi'?

'Mode' would seem to be an easily definable term. Following Jakobson's idea of 'code' (Jakobson 1971), it is often glossed as 'sign system', 'sign-repertoire' or 'semiotic' from which communicators can pick their signs to realise their communicative intentions. Apparently, we intuitively gauge the meaning of 'code' as the quality or type of signs used in a communicative event. Along with 'mode' comes the notion of a 'grammar', i.e. signs belonging to one mode are seen to be governed by a common set of rules that state how these signs can be combined to make meaning in particular situations. In practice, however, things turn out to be less straightforward than this. Let us look at modes a little more closely.

If we take sign quality to mean the different sensory channels via which signs can be perceived and meaning made by communicators, there is justification to subdivide visual, auditory (or audial, cf. Iedema 2003: 32), tactile, olfactory and gustative sign(-mode). While this seems a valid and largely unchallengeable systematisation, it is too rough to be of much use to a theory of multimodality. Language or the linguistic code, for instance, which we justly feel to be a mode in its own right, would thus fall both into the visual and the auditory category as it can be medially realised as either speech or writing. Both are governed by the grammar of language, but as different medial or material varieties of one mode they entail a number of concomitant, additional sub-modes. So speech – besides being linguistic – also employs volume, intonation, timbre, rhythm, speed or pausing, all of which are design features of language in its spoken form and are often termed para-verbal. Furthermore, speech is accompanied and crucially shaped by what has come to be called the non-verbal mode, i.e. gesture, posture and body language. Similarly, writing as the visual counterpart of speech entails typography, which can be seen as the written variant of para-verbal means. What intonation, speed and rhythm are to speech, typography is to writing. However, to complicate matters, some aspects of typography like layout or paper quality are rather non-verbal as they seem further removed from language as such but still accompany writing and contribute to its meaning beyond the linguistic. Moreover, despite being an integral part of writing, the quality of a document's paper transcends the visual mode and has a clear tactile quality to it.

At least three points can be generalized from these brief observations. Firstly, modes cut across sensory channels, so the nature of a sign is not sufficiently characterised by looking at its path of perception. Secondly, one mode can be realised in different media thus creating medial variants of one mode (e.g. speech and writing as variants of the linguistic mode). As any one variant has distinct materiality, it in turn commands individual sets of concomitant sub-modes facilitating or accompanying the variant. This is, among others, a reason why media and

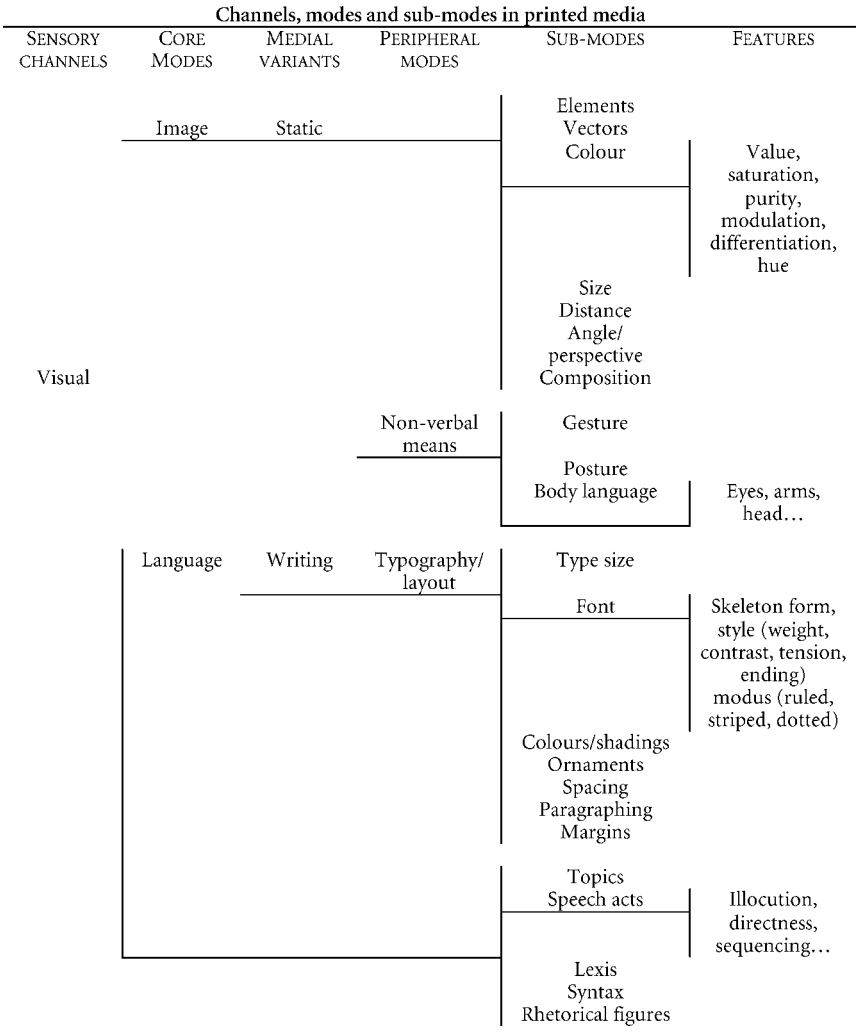


Figure 1. Network of modes, sub-modes and features in printed media

modes ought not be confused but neatly kept apart and regarded in their interdependences. Thirdly, and perhaps most importantly, the range of existing modes represents a hierarchically structured and networked system, in which any one mode can be seen to fall into sub-modes which in their turn consist of distinct features that make up the sub-mode. Let me exemplify these points looking at the image as another major signing mode.

Images, just like language, are not purely visual, they also have a tactile quality to them, which may be reflected in the meaning we construe from them. So while essentially visual, the material and techniques used in the production of the pic-

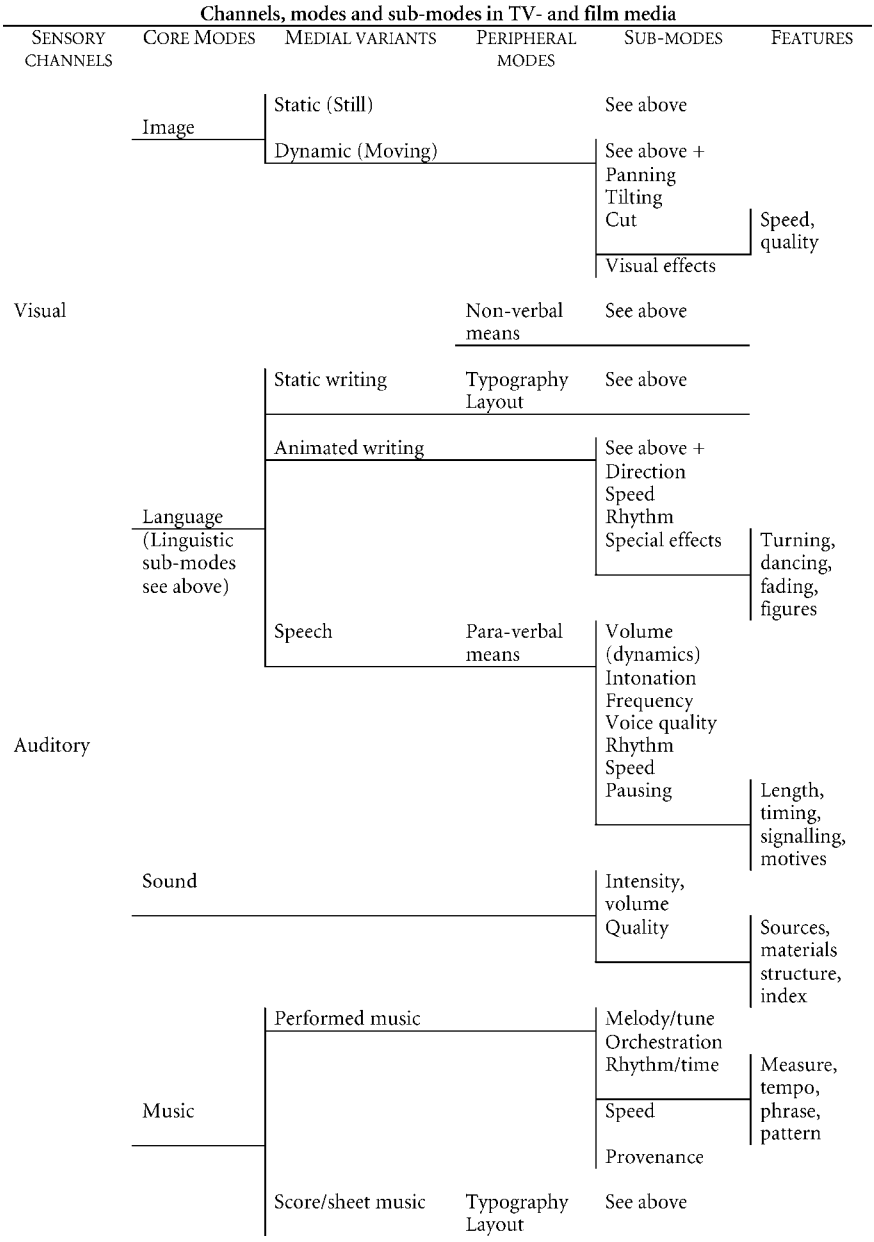


Figure 2. Network of modes, sub-modes and features in TV- and film-media

ture are also subject to touch. The nature of the pictorial sign can, however, not be gauged from its visual quality alone because this also pertains to written language, as we have seen. In this respect both image and language are equally visual.

Images are realised in different media, the static printed image and the dynamic moving image of film and television being the most prominent examples. Besides sharing some sub-modes (e.g. elements, vectors, distance, angle, colour etc.), both medial variants also differ in the set of sub-modes they entail. So the moving image commands such specific sub-modes as, for instance, panning or tilting, it can have narrative[7] and has at its disposal sub-modes originating in post-production like cut or visual effects (e.g. slow-motion, time-laps). Sub-modes in their turn can then be seen to comprise sets of distinctive features, i.e. specific aspects of one sub-mode, which are both phenomena in perception contributing to an overall gestalt as well as analytical categories which help to theoretically come to grips with sub-modes. Colour, for instance, can be decomposed into six distinctive features: value, saturation, purity, modulation, differentiation and hue (Kress & van Leeuwen 2002: 354ff.). Any concrete colour may then be specified by determining values on the scales of those distinctive features and their individual combination.

Applying this kind of systematisation of mode to textual genres in printed media and TV- or film-media, I suggest the following schematically represented network (cf. Figures 1 and 2).

The graphic representations of how modes are structured (cf. Figures 1 and 2) are necessarily formalised simplifications, which demand some comments on inherent problems and limitations.

1. While the columns of core modes contain central sign-repertoires that are deeply entrenched in people's popular perceptions of codes and communication and can stand on their own, peripheral modes come as inevitable 'by-products', as inherent elements of a core mode's specific medial realisation. This is not to say that core modes are more important than peripheral modes or more powerful in terms of their internal grammar and resulting communicative potential. Rather than being a major/minor distinction the differentiation indicates that some modes only come into being along with others and depend on them to some extent. Non-verbal means are ambivalent here as they can also function independently of language. On the one hand, then, the non-verbal is a concomitant aspect of language (cf. Müller 1998), on the other hand, it is part of a communicator's image and thus also relevant to pictorial analysis.
2. In another sense core modes are also abstract modes that need to be instantiated in a specific medial variant. The grammar of language must be realised either in speech or in writing – it is only in these medial variants that peripheral modes pertain to them.
3. Sub-modes[8] constitute a mode in that they provide the building blocks of a mode's grammar. It would be wrong, I believe, to view any of the sub-modes in isolation or as dominant in the make up of a mode. It is rather all sub-modes

in conjunction and their manifold interrelations that establish a mode and facilitate its realisation in a communicative event. As a gestalt in perception, the concrete materialisation of a mode in text and discourse is always more than the sum of its parts. No claim to completeness is made here, both for reasons of space as well as for the incomplete nature of relevant research.

4. If sub-modes are the building blocks of a mode's grammar, then features are the shapes of the blocks and the patterns from which they can be combined. Features can be conceived of as property scales on which certain values can be adjusted in the design of a communicative product/event. Again, it is the sum total of all features and their values that make up the specific realisation of a sub-mode. As they are far from always being binary oppositions, it seems advisable to dispense with the attribute 'distinctive'. For reasons of space I have confined myself to indicating features for only one sub-mode per mode. Within the features listed, again, completeness was not my objective.
5. It is my impression that some sub-modes, like colour, will allow for a relatively finely-grained differentiation into features, whereas others will only command few features or none. Those that consist of a larger variety of features seem to have a more flexible internal grammar and can – due to their complexity – function more autonomously in communication. Such sub-modes will therefore also be able to shift across modes. For instance, colour is not just relevant for pictures but also for typography. Still other sub-modes, like rhythm for instance, are so basic they pertain not just to music, but to speech, sound, animated writing and moving images. Generally, individual sub-modes can apply to more than one mode when the modes in question share some essential design feature, like in the example of rhythm a time-based structure, that is a linear sequencing as its underlying principle. Yet, it would be wrong, I feel, to grant such strong sub-modes the status of modes in their own right as they are always part of a whole configuration of sub-modes, without which the individual sub-mode, however strong, would not be able to operate.[9]
6. Most importantly, the text is the locus where all modes, sub-modes and features are realised. So it is the dynamics of text production and reception, the complex chain from discourse and design to production and distribution (Kress & van Leeuwen 2001: 1–23) that determines how we deploy modal resources and how they in turn are construed in reception. In textual practice one mode is often made salient while others recede into the background – both in perception and production. Genres or text types differ a great deal in how they foreground and background modes and sub-modes. Whereas typography is largely "automatized" (Iedema 2003: 40) in many text types, because it follows established conventions, advertising makes creative use of it and thus puts it centre-stage. Taking a textual perspective on modes and multimodality

means to also pay attention to "the socio-semiotic histories and transitions" (Iedema 2003: 48) of texts, i.e. to uncover how in production and reception modes shift and blend and how producers and users translate or 'transpose' meaning from one mode to the other. In personal e-mail communication or in some types of advertising, for instance, texts are medially realised in writing, while conceptually they are oral in nature. So typically oral textual characteristics normally realised with the help of para-verbal means in speech may be transposed to the written medium using typography and its sub-modes.

3. Mode distinctions

When 'reading' a multimodal text, average recipients will normally become only dimly aware of the fact that they are processing information encoded in different modes. The manifold inter-modal connections that need to be made in order to understand a complex message distributed across various semiotics will go largely unnoticed. All modes, then, have become a single unified gestalt in perception, and it is our neurological and cognitive disposition for multimodal information processing that is responsible for this kind of ease in our handling of multimodal artefacts. A theory of multimodal communication, however, has to meticulously dissect an apparently homogeneous and holistic impression. It has to sensitise us for the essential differences of the modes involved and make us aware of the textual work we invest in building inter-modal relations so crucial to understanding. Multimodal theory also needs to ask in how far there are systematic similarities and ties between the modes involved.

In what follows, I want to demonstrate that it is at least on three different levels that modes can be distinguished from one another: semiotic properties, cognitive orientation, and semantic potential. While I will here confine myself to a juxtaposition of the two core modes language and image,[10] the three theoretical perspectives can be applied to all other modes and sub-modes.

3.1 Semiotic properties

The semiotic properties of a mode refer to its internal structure and to the general ways in which users can make meaning with a mode's signs. Language has what linguists call double articulation, i.e. discrete signs on two levels of organisation, phonemes and morphemes, which combine to form words and utterances. This design feature explains the boundless flexibility and resourcefulness of language. Images, in contrast, have no distinct signing units. There are no rules that would explain how pixels yield higher-level units when combined. What comes closest

to a repertoire (alphabet) of pictorial signs is the elements that can be depicted in images and that we are familiar with from our visual experience of the world. The rules of how to combine them are tantamount to the natural order of things in worlds real or imagined. Consequently then, it is the iconic nature of signs that enables and determines their meaning. Because pictorial signs can access the same mental models that real-world objects access, we can understand pictures. It is the similarity between optical impressions gained from picture viewing and real-world vision that facilitates meaning-making in images.[11] Whereas images thus represent an analogue code, language is deeply symbolic and digital. We need to recode visual perception into abstract sounds or graphs that bear no resemblance to the objects depicted and vice versa. Although writing and pictures share some similarity, as both use the surface area of a medium for representation, they radically differ in that writing utilizes two-dimensional arbitrary graphic forms to represent speech sounds, whereas pictures systematically evoke the three-dimensionality experienced in the perception of objects.

3.2 Cognitive orientation

These and other semiotic properties result in different cognitive operations demanded or afforded by language and images. Most importantly, language is a linear mode that calls for the successive integration of signs into phrases, whereas images are rather based on simultaneous and holistic gestalt-perception. Consequently, images can be regarded as a quick mode relative to language as they do not necessitate parsing. We know from psychological experiments that images are far more likely to be attention-getters in perception than language and can also be memorised much more easily and effectively. Both have to do with their analogue code characteristics – no recoding needs to take place and pictures can therefore be regarded as a code close to reality or – as some semioticians have argued – a 'language' without a code. The speed of pictorial perception is usually put down to the simultaneity of gestalt formation, whereas the communicative impact of images is seen in the fact that they directly tap into the emotions and provide immediate sensory input.

3.3 Semantic potential

Semiotic and cognitive characteristics determine what users of a mode can do with it in terms of specific meaning-making resources. Although debatable, it has generally been accepted that the semantics of language is less vague and polysemous than that of the image. While language provides scope for double meaning, it has conventional semantics attached to words and utterances. Images, on the contrary,

are seen to be inherently vague and ambiguous and can only be made to mean and communicate specific contents by a combination with other modes or the embedding into narrowly defined communicative situations. Most importantly, images lack a definite speech act repertoire, which is why their illocutions remain cloudy unless they are complemented by language. Language, on the other hand, counts as less rich in information than images, which carry a welter of sensory information and are particularly intense in terms of connotation. Conversely, language is at a great advantage as for its potential to communicate all sensory modalities, whereas pictures clearly are confined to visual information.[12] Similarly, the self-referential capabilities of images are weak, whereas they are basically unlimited with language. Finally, language can be used to make just about any utterance imaginable. This huge semantic flexibility, which results from the linguistic design principle of double articulation and an elaborate set of rules, is contrasted by some obvious semantic restrictions of images. Some meaning relations like causality cannot be expressed, negation and affirmation are impossible and the utterances construable from images are usually additive. To sum up, language has its strength in the depiction of events and states-of-affairs in time, whereas images are particularly suited to the representation of objects in space and their physical characteristics (cf. Kress 1998: 68f.).

4. Mode integration: Overlapping and mixing

The comparison between language and image showed up a number of essential differences, which shape the ways these two central signing modes can be made use of in communication. The danger inherent in contrasting two modes, however, is that we tend to somehow look at one mode in terms of another. So, mostly, due to language's dominance, we seem to be asking which linguistic properties images have. Thus we run the risk of overlooking some important design features of images which are outside the linguistic perspective. While a possible denigration of images resulting from this must be avoided, the metaphoric stance of a pictorial language is engrained in our naturally logo-centric take of communication.

More importantly, we need to realise that the mode distinctions pointed out are relative, as language and image are inseparably intertwined both in concrete forms of communication (mode mixing) as well as cognitively, semantically and historically (mode overlaps). Producing and understanding images presuppose verbal categorization of visual elements just as producing and understanding language relies heavily on conjuring up and manipulating mental images, the latter being most poignantly illustrated by the vast store of phrases that use a literal image to make metaphoric meaning. The history of alphabetic signs is a testimony

to the fact that the graphic articulation used to represent speech sounds has its origins and immediate precursors in the pictorial. The strongest argument for the innate tie between language and image, after all, is their co-presence in almost all forms of communication, a symbiotic mode integration (mixing) which is guided by the principle of reciprocally balancing out limitations and weaknesses of the modes combined.

5. The textual interface between language and image

There are two basic ways in which the linguistic and the pictorial mode can come together in a text. Firstly, a verbal text can itself acquire image qualities by means of typography and layout. In this case a peripheral mode (typography) of a medial variant (writing/language) is employed for a partial transfer from one core mode (language) to another (image). Here, the carrier of the linguistic mode emulates the pictorial. Secondly, and this is the more common option, a verbal text is combined with an image. The two core modes are semantically and formally integrated so that each mode strategically employs its range of sub-modes thus unfolding the specific semiotic potential of each mode and contributing to an overall communicative gestalt. A specific type of this language-image-combination would result if the verbal text contained language that was itself pictorial or figurative and established a semantic or formal link with the accompanying visual image. Let me illustrate both types of language-image-link by discussing two sample texts drawn from the advertising genre. My aim in this will be to show on which levels such interfaces between two modes can be analysed.

5.1 The typographic image

In an advertisement of the RSPCA (the Royal Society for the Prevention of Cruelty to Animals) for free-range eggs (cf. Figure 3) the verbal text is typographically designed to yield the visual form and appearance of a supermarket receipt. Although, of course, the language contained in the text is not what we would expect to read on a receipt, conjuring up the image of a receipt is possible, because this specific text type, like many others, comes with a built-in range of graphic features that can be imitated. Wehde (2000) calls such configurations of typographical and layout properties which form a set of visual expectations tied to a particular text type (format) a "typographisches Dispositiv", which could be rendered in English as the 'typographic repertoire' of a text type. Which typographic/layout sub-modes, then, have been employed to give the impression of the typographic repertoire of a receipt?

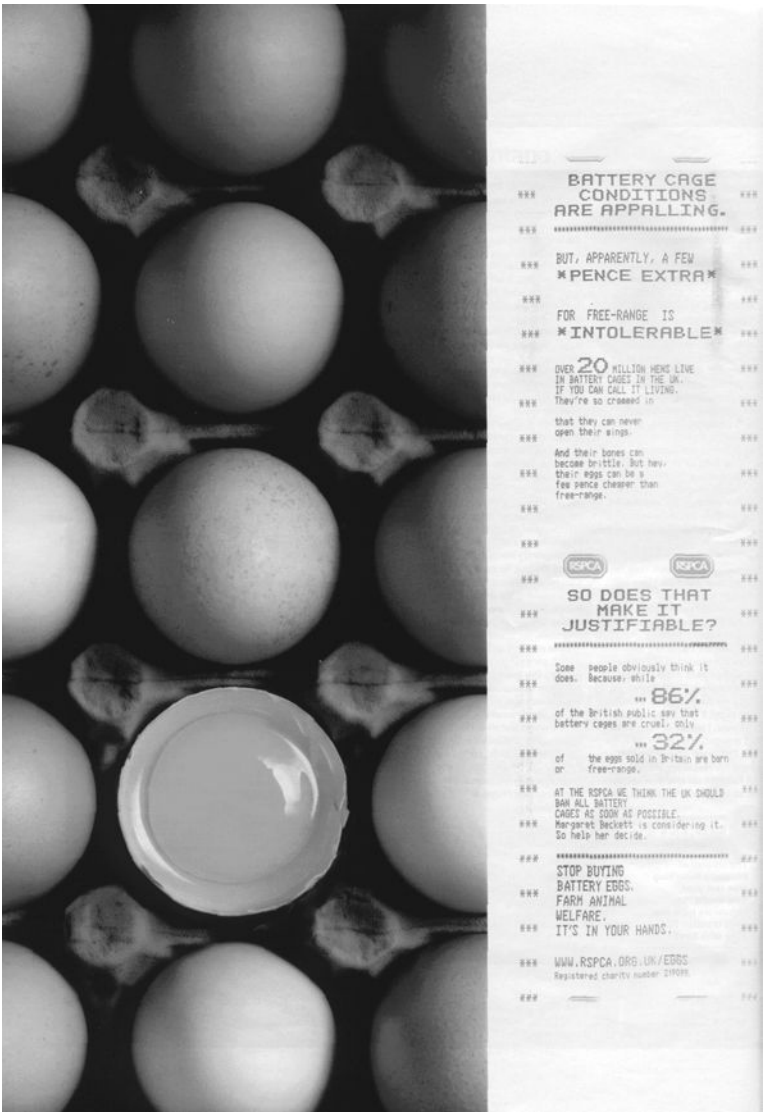


Figure 3. "A few pence extra", RSPCA (The Times Magazine 15.03.03, p. 10)

Battery conditions are appalling. But apparently, a few pence extra for free-range is intolerable. Over 20 million hens live in battery cages in the UK. If you can call it living. They're so crammed in that they can never even open their wings. And their bones can become brittle. But hey, their eggs can be a few pence cheaper than free-range. *So does that make it justifiable?* Some people obviously think it does. Because, while 86% of the British public say that battery cages are cruel, only 32% of the eggs sold in Britain are barn or free-range. *At the RSPCA we think the UK*

should ban all battery cages as soon as possible. Margaret Beckett is considering it. So help her decide. *Stop buying battery eggs. Farm animal welfare. It's in your hands*. www.rspca.org.uk/eggs Registered charity number 219099

The visual image of a receipt is mainly formed by narrow margins, which have been marked by lines of three stars each. The text body is heavily subdivided into small portions of variously aligned print. This has been achieved by paragraphing and spacing as well as by lines formed of the same stars as at the margins. The font clearly betrays its provenance as having been produced by the typical cash-desk printer. This is signalled through the formation of the characters from individual dots, a system also used in digital displays. Contributing to this is the blue colour of the print as well as its irregular quality and the blurred print blotches in between, which by association also indicate the low but functional quality of the paper and the printing technology. The enumerative and strongly portioned character of the receipt also materialises through bold print, tab stops in the middle of lines, capitalisation and the use of numbers so typical of receipts. Besides being heavily paragraphed, receipts also come as parts of a continuous text (paper roll) from which they are torn when handed out to the customer. This continuous character of the text is borne out by the recurring logo on either side and the centrally aligned text body right underneath, which usually communicates the name and address of the supermarket or some other standing detail. Continuity is also expressed by the cut off logos at the top and bottom of the text.

When typographical repertoires are exported from one text type (receipt) to another (advertisement), the resulting effect is not merely pictorial as in our example, where the receipt is sort of reified as a textual object. More importantly, exported or emulated typographical repertoires have a semantic impact. In the RSPCA advertisement the receipt-like character of the text adds to the meaning of the verbal text. It supports the central argument of the ad, which says that purchasing behaviour can make all the difference in the battle for more free-range farming. The receipt image of the text makes the pivotal point that it is in the supermarket where farming policies are shaped via the price of the eggs and consumer behaviour. The readiness to overcome one's own meanness and spend a few pence extra on eggs as the target of the advertisement is suitably transported in the visual image of a receipt, as it directly taps into the knowledge script of shopping and thus reduces the mediated effect of communication and makes it more direct.

5.2 The verbal image and the language-image-link

Print advertising is a textual genre whose reliance on language-image-combinations is almost obligatory. In the design of such inter-modal links creators of ads exploit the whole range of semantic, formal and pragmatic relations, which can be

Figure 4. "No small fry", Toshiba (The Sunday Times Magazine 14.09.97, pp. 2f.)

anything from simple to complex.[13] Linguistic text analysis, however, is invariably confronted with problems when it is to tackle such two-mode genres. Let me, therefore, suggest some levels and criteria of analysis suitable to shed light on the interface between language and image.

An advertisement for Toshiba laptops (cf. Figure 4) combines an image with a long copy, whose main feature is the accumulation of a number of metaphoric phrases (bold), which lend salience to the verbal text structure. Taken together, the figurative expressions and their core lexical elements (e.g. *ocean, waves, shoal, swim, plunge; catch, small fry, haul, bait, canned*) help to form two contiguous mental images or knowledge frames: the sea and fishing. The visual image (a half-open tin showing a keyboard inside and some computer features on the outside, like CD-compartment or connections for periphery devices) helps to focus the mental image of fishing and fish processing, but also links it to the main subject of the text, portable computing. The pivotal question we need to ask is how the two modes can come together in the make-up of the text and which pragma-semantic and cognitive operations underlie the meaning-making processes in this specific instance.

> We've squeezed a desktop into a portable. *No catch* (1). The only difference between a Toshiba portable and a desktop is scale. When it comes to size and weight the two are *oceans apart* (2). And for power, function and features our portables are *no small fry* (3). Features like the best chips, Intel MMX, the largest screens, 13,3" and the fastest CD-ROM drives, mean our portables are *a good haul* (4). And the PC Card Slots on every Toshiba mean our portables can *take the bait of most networks* (5). Whilst our innovative firsts have always *created waves* (6) amongst desktop and portable manufacturers alike. It's because we would *rather lead the shoal than swim with it* (7), that makes us the world leader in portable computing. Our portables have consistently *landed prizes* from PC Magazine for service, reliability and technical innovation. If you'd like to *take the plunge* (8) and

> see why more people are working wherever and whenever they like, call us on ... Hopefully you'll agree, Toshiba really do *have portable computing canned* (9)...

The first level to scrutinize is the semantic and pragmatic tie between language and image. It is obvious here that both the image and a whole network of linguistic expressions function metaphorically. The central metaphor could be spelt out as LAPTOP = FISH CAN (visual image) or BUILDING LAPTOPS = CATCHING FISH/CANNING FISH (verbal text). Whereas the visual image provides the overriding metaphor which serves as a frame of orientation for the interpretation of the verbal metaphors, the literal meanings of the metaphorical phrases (in italics) detail and structure the emerging mental images (sea, fishing). The result of this kind of visual-verbal metaphorical play is the mental mapping of source domain features (sea, fishing) onto target domain features (computer manufacturing). So here analogies are built between

(1) the sea and the market (*oceans apart, create waves, take the plunge*)

(2) the quality/quantity of the catch and the quality of the computers (*no small fry, good haul*)

(3) fish processing (canning) and quality computer assembly (*have portable computing canned*)

(4) fish/fishing and computer firms (*rather lead the shoal than swim with it, no catch*)

(5) instruments of fishing (bait) and compatibility with networking standards (*take the bait from most networks*).

On a second level, we could ask which cognitive operations are afforded by the design of the language-image-link. Clearly, what is intended here is the oscillation between literal and metaphoric meaning-making throughout the text, but also the successive integration of verbal phrases into a whole. On the one hand, all figurative expressions can be read literally (i.e. with recourse to sea and fishing), an interpretation facilitated mainly by the visual image, which provides the concept of a can with all its physical properties. On the other hand, of course, the phrases can be decoded in their metaphorical meanings, which support the persuasive intention of the text. Both on the literal and metaphorical plane of meaning the phrases combine to build a structured network. The network's elements cohere because on a literal level they build paradigmatic sense relations and all add to a common mental image. On a metaphorical level the phrases are networked as they all contribute to realising typical advertising speech acts and establishing the advertising text pattern. So, *no small fry* and *good haul* describe and evaluate the advertised product, *created waves, rather lead the shoal than swim with it* and *have portable*

computing canned promote the firm's image and express praise, whereas *take the plunge* is an appeal to the consumer to try out the advertised products.

On a third level of analysis, the overall textual structure built from language and image in our sample text is – as we have seen – one of metaphorical projection and literalisation. Verbal and pictorial text are strongly interdependent in as much as the visual image promotes the literal readings of the figurative phrases, and the metaphorical language explains the context and the motivation for the pictorial metaphor.

Finally, on a fourth level of analysis we need to enquire into the specificity of the visual image. Images in concrete communicative events always come as types, and there are a number of design features in our example that are for one thing typical of advertising and, for another, facilitate the language-image-link described so far. Most importantly, the image's metaphoric nature is realised by a morphing technique which allows for the carefully engineered blending of visual features. As a result the reader gets a very realistic impression of an imaginary object, which plays tricks on his perception. Shape, size and colour come as unifying characteristics of both objects (can, laptop) morphed into a single gestalt, while can opener and keyboard as well as the plug-in connections represent distinguishing traits of the objects blended. In advertising images single objects are often shown against a neutral background in order to bring out their salient characteristics, as has been done here. Also, there is something like a functional perspective in advertising images, which makes the perception of the objects depicted as easy and effective as possible. The can/laptop is shown from slightly on high and from a relatively short distance so we can easily take in all its important attributes. Curiosity and tension is created by the half-open state of the can with the opened part pointing away from the viewer.

6. Semiotic principles across modes

The two sample texts were to demonstrate that mode-integration may be complex, because there is both mode mixing, i.e. the calculated and complementary co-deployment of language and image, as well as mode overlapping, i.e. the collapsing of modes into one another. Overlaps of modes are seen in instances where language can be doubly pictorial. Firstly, verbal text can assume pictorial quality via typography and layout and secondly, language can be based on and evoke mental imagery. Mode-overlapping, however, also reflects in the easily neglected fact that images are to a great deal rooted in language or rather in knowledge frames and scripts which are heavily codified and structured in the form of our linguistic repertoires. So, what can be expressed and communicated in images (in produc-

tion and reception) is not only dependent on our visual experience of the world or the material and technical properties of image-media, but is also crucially shaped by our stock of words, phrases and stereotypical language utterances.

When separate modes indeed so closely intertwine in multimodal texts, is it not likely, then, that "common semiotic principles operate in and across different modes" (Kress & van Leeuwen 2001:2)? This is a central question of multimodal theory and analysis. As we have seen in the case of language and image (cf. Section 3), modes differ noticeably from one another in terms of internal sign structure, semantic potential and cognitive operations afforded. This is why individual modes need to first of all be regarded as possessing their own 'grammars', which are distinct from one another as they follow different organising principles and make different functionalities available. This does not, however, rule out the possibility that some overriding principles govern and guide all modes simultaneously. Such trans-modal operating principles, then, would have to be sufficiently general and basic to be able to span the great variability of modes. My view on common semiotic principles across modes, therefore, is a balanced and dynamic one. I endorse the formal and functional differentiation between modes while at the same time acknowledging the trans-modal operation of very global semiotic rules governing the organisation of individual modes and their reciprocal integration. The general cross-modal principles are only instantiated in texts and communication, and I suggest to look at those principles as a means to regulate and guarantee a kind of semiotic equilibrium in a concrete multimodal text. In what follows I will roughly sketch out some of the basic semiotic principles operating across modes.

1. The three Hallidayan *meta-functions* (Halliday 1994) would be the first principle that can easily apply to all modes imaginable and to the multimodal text as a whole. Any mode is – to varying degrees – able to depict states-of-affairs (ideational), design some social interaction between the communicators (inter-personal) and contribute to organising and structuring the text (textual). In any one multimodal text these three functions need to be fulfilled and, more importantly, distributed across the modes present. Here, the aim must be an inter-modal balance between the meta-functions, that is their distribution across modes will be guided by how, in a given communicative event, the functions can be realised most efficiently. Consequently, modes will be positioned towards one another according to which part they play in fulfilling the meta-functions. This is a first key to the structuring of multimodal discourse.
2. *Segmentation*, that is the decomposability of larger sign structures or gestalts of perception into their constituent elements, would seem to be a second semiotic principle operating in and across modes. All modes need to signal their internal structure as keys to the retrieval of portions or layers of meaning.

But segmentation also applies to the organisation of the multimodal text as a whole, which needs to be indicated by structurally salient signs connecting modes and specifying their semantic ties. *Syntax*, then, in its widest sense is a core feature of any accumulation of signs, both with respect to the internal structure of one mode as well as with respect to the overall structure of a multimodal text.

3. Meaning in texts comes on three interrelated planes. Signs can refer to concepts (*denotation*), they can convey concomitant, socially shared emotive or evaluative meaning (*connotation*), and signs can also activate and tap into purely individually valid facets of meaning (*association*). Modes seem to differ in their ability to provide for those three types of meaning: music, for instance, seems weak on denotative meaning, but strong on associative meaning. Language, on the other hand, has its strength in the domain of the denotative, but can also service other layers of meaning. However, it has relatively little potential on the associative plane. Again it is my claim, here, that multimodal texts need to generate and conserve a kind of balance between denotation on the one hand and connotation/association on the other. The deployment of modes in a multimodal text will seek this semantic equilibrium, and hence its structure will reflect the adherence to this inter-modal principle. So in a commercial, for instance, sound and language will usually be geared towards the denotative, whereas music clearly provides associative meaning. The image stands midway between the two as it can deliver both denotations as well as associations.
4. Another common semiotic principle of multimodal communication can be seen in the necessity of modes to build *semantic relations* to concomitant modes. As soon as two modes are combined they will automatically create *semantic ties*, which can be of a great variety. Meaning made in one mode may be repeated or paralleled in another, or it may be complemented, negated, contradicted or reinterpreted etc. Inevitably, co-deploying modes will cause them to come together in perception, and this simultaneity to the senses must be construed semantically in some way or in another.
5. Signs can make meaning in three ways. As *icons* they can, by virtue of their similarity with the depicted objects, simulate real or possible worlds. *Indexical signs* point to objects or states-of-affairs other than the one denoted by way of logical operations and mental contiguities. Finally, signs can be *symbols* arbitrarily standing for some meaning not otherwise connected to the signs. These three signing options are universal meaning-making principles that can apply to all modes, although, of course, modes are known to favour one or the other way of meaning-making. Language is a dominantly symbolic mode, whereas the image is predominantly iconic in nature, yet it may well act as symbol

or index, while language may be used iconically or indexically. Again, using modes or individual sign complexes from one mode as icon, index or symbol and thus relating modes to one another is a major structuring device in multimodal communication.

6. Finally, I believe that recognizing *gestalt similarities*, i.e. inter-modal analogies on both the formal and semantic level, is the core mental operation required in multimodal communication. This applies to the perspective of the producers, who craft *analogies*, just as well as to that of the recipients, who have to recover those analogies and interpret them. Analogies are so central to the design and understanding of multimodal discourse because it is mainly thanks to them that individual signs from various modes cohere into a common whole and form a unified gestalt in perception (Stafford 2001).

7. Conclusion

My observations on multimodality were focused on the interface between language and image. More than anything else, it should have become obvious that inter-modal relations like those in the language-image-link are anything but simple and pose both theoretical and analytical problems. For one thing, the notion of 'mode' justifies the view that what we do in multimodal communication is, toolkit-like, pick from a set of signing resources according to concrete communicative intentions and an envisaged overall design of the text. For another, this perspective is too static. While it is certainly true that modes have their individual characteristics (semiotically, semantically and cognitively) which pre-determine how they can be deployed in a textual structure, the dynamics of meaning-making must be given due emphasis. In the production of multimodal texts modes and their respective sub-modes blend, shift and mix in possibly myriad ways. It is only through a multi-faceted and integrated analysis on all levels of text, which is susceptible to the dynamic processes of "inter-semiotic shifts" (Iedema 2003:42), that meaning-making can be reconstructed. The meaning recovered in such analyses is both a reconstruction of meaning deployed by the producers as well as an interpretation of the meaning construable by recipients.

Notes

1. The confusion between multimediality and multimodality can be seen as an effect of the hype surrounding the discipline, which has mainly been generated by a fascination with new technologies and their apparently boundless opportunities.

2. Iedema (2003: 40) calls this kind of argument the "always already (*Überhaupt*) multi-semiotic nature of meaning-making".

3. For a short historical sketch of multimodality's comparatively brief history (see Iedema 2003: 30ff. and Stöckl 2004: 11–20).

4. Such 'grammars' of individual modes are, for instance, outlined in Kress and van Leeuwen (1996), Doelker (1997) or Stöckl (2004) (image), in van Leeuwen (1999) (sound and music), in O'Toole (1994) (displayed art, e.g. sculpture), in Stötzner (2003), Willberg and Forssmann (1999), Walker (2001) or Wehde (2000) (typography), and in Kühn (2002) (non-verbal communication). While some of them are based on an explicitly systemic functional approach, others apply different methodologies. Common to most theories of single signing modes is the (metaphorical) transfer of some kind of linguistic or semiotic pattern. This goes to show that language is seen to be central in signing practice.

5. Kress (1998: 55ff.), for instance, suggests that it is first and foremost a change in social conditions, and not technology as such, which drives the shift from language to image in the communicative landscape. He pinpoints information overload as a cause for a greater reliance on the image.

6. Kress and van Leeuwen (2001) is the first attempt to outline a general theory of multimodal communication. The objective to be all-embracing has, however, been pursued, here, to the detriment of specificity.

7. Stöckl (2002) shows how the static printed image is also capable of realising narrative.

8. The terminology used for sub-modes and features here is heterogeneous. I have largely followed Kress and van Leeuwen (1996) and van Leeuwen (1999), but have also used rough labels borrowed from generally accepted vocabulary. As for typography, I have made use of Stötzner (2003) and Willberg and Forssmann (1999). For inspiration, I have also turned towards Neuenschwander (2001) and Bellantoni and Woolman (1999). As for the para-verbal, I have consulted Neuber (2002). The linguistic sub-modes follow accepted notions of levels of text-analysis. Again it needs to be emphasised that my aim was a very general but systematic scheme of things, not total completeness or precision. An application of multimodal analysis to the TV-commercial which roughly follows the lines sketched out here can be found in Stöckl (2003).

9. Kress and van Leeuwen (2002: 350f.) can be understood to generally endorse the idea that colour is a mode in its own right, although they also advise caution. Their main argument for regarding colour as a mode is that "it can combine freely with many other modes" (ibid.: 351). Although that seems true, colour, when combining with other modes, is part and parcel of those modes, which is why I would like to maintain that colour comes as a sub-mode.

10. The following differences between language and image are presented in more detail and with relevant bibliographical sources in Stöckl (2004: 245ff.). Almost all of the arguments represent a broad consensus in semiotics and cognitive psychology.

11. A concept of images based on their iconic nature can also be maintained when applied to abstract, i.e. non-depicting images. Colours and shapes in spatial combinations that do not refer to concrete objects real or imagined will mainly make 'meaning' by associations we have with these colours and shapes acquired in sensory experience.

12. By synaesthetic connections images can also communicate information other than the purely visual.

13. Stöckl (2004) is a detailed study of the language-image-link with respect to advertising and journalism.

References

Bellantoni, Jeff & Woolman, Matt (1999). *Type in Motion. Innovative digitale Gestaltung.* Mainz: Hermann Schmidt.

Doelker, Christian (1997). *Ein Bild ist mehr als ein Bild. Visuelle Kompetenz in der Multimedia-Gesellschaft*. Stuttgart: Klett-Cotta.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar.* London: Arnold.

Iedema, Rick (2003). "Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice." *Visual Communication, 2* (1), 29–57.

Jakobson, Roman (1971). *Fundamentals of Language*. The Hague: Mouton.

Kress, Gunther (1998). "Visual and verbal modes of representation in electronically mediated communication: The potentials of new forms of text." In I. Snyder (Ed.), *Page to Screen. Taking Literacy into the Electronic Era* (pp. 53–79). London: Routledge.

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images. The Grammar of Visual Design*. London: Routledge.

Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold.

Kress, Gunther & van Leeuwen, Theo (2002). "Colour as a semiotic mode: notes for a grammar of colour." *Visual Communication, 1* (3), 343–368.

Kühn, Christine (2002). *Körper – Sprache: Elemente einer sprachwissenschaftlichen Explikation non-verbaler Kommunikation.* Frankfurt am Main: Peter Lang.

Müller, Cornelia (1998). "Beredte Hände. Theorie und Sprachvergleich redebegleitender Gesten." In C. Schmauser & T. Noll (Eds.), *Körperbewegungen und ihre Bedeutungen* (pp. 21–44). Berlin: Arno Spitz.

Neuber, Baldur (2002). *Prosodische Formen in Funktion. Leistungen der Suprasegmentalia für das Verstehen, Behalten und die Bedeutungs(re)konstruktion*. Frankfurt am Main: Peter Lang.

Neuenschwander, Brody (2001). *Letterwork. Creative Letterforms in Graphic Design.* London and New York: Phaidon.

O'Toole, Michael (1994). *The Language of Displayed Art.* London: Leicester University Press.

Stafford, Barbara Maria (2001). *Visual Analogy. Consciousness as the Art of Connecting*. Cambridge, MA: MIT Press.

Stöckl, Hartmut (2002). "From space to time into narration – Cognitive and semiotic perspectives on the narrative potential of visually structured text." In H. Drescher, W. Thiele, & C. Todenhagen (Eds.), *Investigations into Narrative Structures* (pp. 73–98). Frankfurt am Main: Peter Lang.

Stöckl, Hartmut (2003). "'Imagine': Stilanalyse multimodal – am Beispiel des TV-Werbespots." In I. Barz, G. Lerchner, & M. Schröder (Eds.), *Sprachstil – Zugänge und Anwendungen. Ulla Fix zum 60. Geburtstag* (pp. 305–323). Heidelberg: Universitätsverlag Winter.

Stöckl, Hartmut (2004). *Die Sprache im Bild – das Bild in der Sprache. Zur Verknüpfung von Sprache und Bild im massenmedialen Text: Konzepte, Theorien, Analysemethoden*. Berlin: de Gruyter.

Stötzner, Andreas (2003). "Signography as a subject in its own right." *Visual Communication, 2* (3), 285–302.

van Leeuwen, Theo (1999). *Speech, Music, Sound.* London: Macmillan.

Walker, Sue (2001). *Typography and Language in Everyday Life. Prescriptions and Practices.* London: Longman.

Wehde, Susanne (2000). *Typographische Kultur: eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung.* Tübingen: Niemeyer.

Willberg, Hans-Peter & Forssmann, Friedrich (1999). *Erste Hilfe Typographie. Ratgeber für Gestaltung mit Schrift.* Mainz: Hermann Schmidt.

Chapter 2

Modelling multiple semiotic systems

The case of gesture and speech*

Peter Muntigl
University of Salzburg, Austria

In systemic functional linguistics, language (i.e. speech and writing) is often taken as the central semiotic system through which we construe experience as meaning and enact social reality. In addition, other semiotic systems such as image and gesture are considered to be modelled in the image of language; that is, they are metafunctionally and stratally organised in the same way that language is. From this premise, it becomes necessary to investigate the kinds of metafunctions that other semiotic systems may deploy and the degree to which they are stratally organised. Additionally, we might ask the question whether or not semiotic systems such as gesture and image have a grammar? Drawing mainly from the relationships between gesture and speech, this chapter suggests that semiotic systems should be viewed along a continuum between proto-language and language, and, ideationally, that multiple semiotic systems deployed in meaning-making are functionally organised and interrelated via transphenomenal categories of expansion and projection. Finally, it will be proposed that the contextual variable *mode* works as the primary organiser that is responsible for activating combinations of semiotic systems in contexts of situation.

1. Introduction

The observation that people draw on various meaning-making systems when construing experience or enacting social reality is certainly not new. Research that may be loosely subsumed under the rubric of 'context analysis' has shown how various 'semiotic systems' such as *speech*, *gesture*, *body position* and *eye gaze* are simultaneously deployed in interaction (for an overview, see Kendon 1990: 15–49). The interactive coordination of various modes has already been documented in terms of the relationship between gesture and speech (Kendon 1982, 1997; McNeill 1992; Streeck 1993, 1994), speech and gaze (Goodwin 1979; Kendon 1990; Streeck 1993), speech and body position (Kendon 1985) and speech, gaze, gesture and

body position (Goodwin & Goodwin 1992). A more recent focus has also shown how technologies and/or spatial arrangements in workplace or organisational settings play a part in shaping interaction (Goodwin 1994; Goodwin & Goodwin 1996; Hutchins 1996; LeBaron & Streeck 1997).

The meaning-making role of semiotic systems other than speech has also been addressed in systemic functional linguistic (hereafter SFL) oriented research. Kress and van Leeuwen (1996), for instance, have argued that images have a grammar, and may be analysed along the same lines as Halliday's (1978) metafunctional approach to language. Music has also been given a semiotic interpretation (van Leeuwen 1999). Furthermore, Kress and van Leeuwen (2001) claim that smell and colour may be treated as semiotic systems in their own right, and may be added to the list of systems used in meaning-making.

In examinations of how meaning is created, SFL research has been largely selective, favouring speech or writing as the primary semiotic systems through which experience is construed as meaning.[1] In this chapter, I discuss how other semiotic systems – with a special focus on *gesture* – may be included alongside speech in the construction of meaning. The first question I address involves how semiotic systems can or should be modelled; that is, how do semiotic systems such as image or gesture relate to 'language' (i.e. speech or writing), and can these semiotic systems be modelled using what Matthiessen (1993: 230) refers to as "general principles of intra-stratal organisation"or "fractal dimensions". My second question concerns how the deployment of semiotic systems may be organised. Put differently, can a situational variable such as *mode* be made responsible for coordinating which semiotic systems may be activated in various contexts of situation?

2. Language vs. semiotic systems

As semiotic systems speech and writing are given very high status in SFL. First, Halliday and Matthiessen (1999) argue that the semiotic systems that best fall under the category of 'language' are speech and writing; and second, language (i.e. speech and writing) is the primary semiotic system through which people construe experience as meaning. They also suggest that other semiotic systems such as music, dance, dress, cooking, the organisation of space, charts, maps and diagrams are modelled on language:

> These systems enter into relations with language in two ways. On the one hand, they are metonymic to language: they are complementary, non-linguistic resources whereby higher-level systems may be realised (e.g. ideological formations realised through forms of art; theoretical constructs realised through figures and diagrams). On the other hand, they relate metaphorically to language: they are

> constructed, stratally and metafunctionally, in the image of language itself, and hence can be modelled on language as prototype, being described "as if" they had their own grammar and semantics. (Halliday & Matthiessen 1999: 606)

The question as to whether other semiotic systems are metaphorical versions of language is an important one. One reason is the question of modelling different semiotic systems in the meaning-making process. If Halliday and Matthiessen are right, then we would want to model other semiotic systems in terms of metafunctions; that is, they should be metafunctionally modelled along the same lines that language is. But, before I pursue the question of how other semiotic systems should be modelled, it is important to reflect on why other semiotic systems are not considered languages in SFL.

What is a language? In SFL there are a number of perspectives that need to be taken into account to answer this question. First of all, Halliday (1978) argues that languages are tri-stratally organised in terms of content – semantics and lexicogrammar – and expression – phonology/ graphology (represented in Figure 1). In this way, Halliday differentiates between what he calls proto-language – a semiotic system containing only two strata, consisting of semantics and phonology – and a language. Infant speech is characteristically a protolanguage. In the early stages of ontogenesis an infant's linguistic system consists of a limited set of meanings that are sounded in various ways (Halliday 1978). In the transition to language children develop a lexicogrammar that allows them to 'mean' in a vastly increased number of ways. It is what Halliday refers to as "the explosion into grammar"that marks the early stage of language development. Language, in addition, exhibits an interstratal organisation in which meaning is realised in the realisation of wording in sound (for a discussion of *realisation*, see Halliday 1992).

Seen in terms of instantiation, a language is also a meaning potential. Each of the strata – semantics, lexicogrammar and phonology – is a resource of meaning-making that is represented paradigmatically by system networks. Minimally then, in order to be a language, a semiotic system must be tri-stratally organised in terms of semantics, lexicogrammar and 'expression', and its strata must be paradigmatically organised as systems of meaningful choices. What I have just outlined is, arguably, a synoptic perspective on language. This perspective does not take ontogenesis into account. One could interpret Halliday's description of proto-language and language as a continuum in which semiotic systems begin as simple content-expression systems and develop into tri-stratal meaning systems that incorporate a grammar.

The proto-language/language continuum may also exist for other semiotic systems. An example is sign language. As a language, sign exhibits all the characteristics of speech. There are of course differences. Whereas speech has phonology as its expression plane, sign has *kinesic* structure and includes use of arms, body

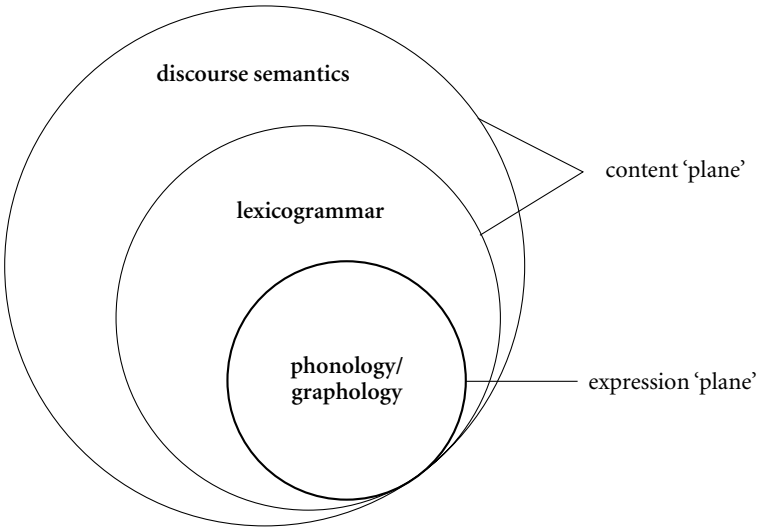


Figure 1. Tri-stratal organization of language (from Halliday & Martin 1993: 26)

Gestures	Pantomime	Emblem	Sign Language
←			→
1. obligatory presence of speech			obligatory absence of speech
2. linguistic properties absent			linguistic properties present
3. not conventionalised			fully conventionalised
4. global & synthetic			segmented and analytic

Figure 2. Kendon's continuum (adapted from McNeill 2001: 2)

posture and head movement. Put another way, speech is marked by meaningful acoustic-temporal patterning, and sign by meaningful visual-spatial (and temporal) patterning. Sign should, however, not be confused with gesture, and it is worthwhile pointing out the differences between these two semiotic systems. McNeill (1992, 2001), for instance, drawing from Kendon's (1988) work, argues that gesture and sign language may be placed along what he now refers to as 'Kendon's continuum', shown in Figure 2. Gestures and sign language are located at opposite ends of the continuum, with pantomime and emblems located at points further towards the centre. According to McNeil gestures obligatorily accompany speech, have no linguistic properties, are not conventionalised and are global and synthetic. In contrast, signing occurs in the absence of speech, exhibits linguistic properties, is fully conventionalised and is segmented and analytic.

As McNeill (2001) points out, however, there is nothing preventing gestures from acquiring language-like properties. Deaf children of hearing parents, for instance, have been found to communicate with a gestural system that manifests a lexicon and basic syntax (Goldin-Meadow & Mylander 1984). In this way then, Kendon's Continuum may be viewed ontogenetically, in that gestures could be seen as the proto-languages of sign languages. One important difference, however, is that gestures do not develop into full-fledged languages in the non-deaf.

In SFL it seems that tri-stratal and paradigmatic organisation would not be sufficient enough to elevate a semiotic system to the status of a language. This is because there still remains the problem of a semiotic system's degree of independence from speech. If a semiotic system other than speech could be deployed in a variety of situations, then it could be argued that the semiotic system in question is functioning as a language. Following McNeill, gestures would be ruled out, since they only occur in the presence of speech and cannot fully take over the functions of speech. Perhaps it is the case that if a semiotic system has developed into a language, then other semiotic systems will not compete with it. Once you have got a language, it will be the favoured semiotic system to be activated in a context of situation. This does not mean that other semiotic systems will not be activated, that they cannot add meaning or that they cannot displace language in various situations. All of these can and do happen. Nevertheless, language or not, if we want to understand what different semiotic systems are doing, we must examine to what degree they are stratally organised and the range of meaningful options they make available in situations of meaning-making.

2.1 Stratal organisation and grammar

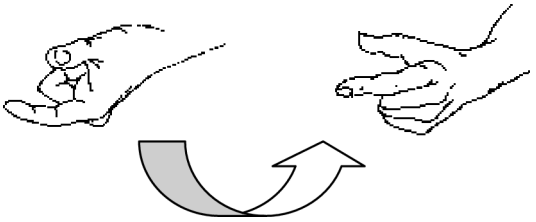
Returning to Halliday and Matthiessen's (1999) claim that various semiotic systems may be described *as if* they had their own grammar, we are still left wondering whether, genuine or not, they actually have a grammar. And, if they do, what the grammar looks like. Recall from McNeill's standpoint, that gestures do not exhibit linguistic properties. This does not mean that gestures are not meaningful. Quite to the contrary, McNeill argues quite pointedly that gestures are non-redundant and that they 'add' meaning to the speech segments with which they co-occur. By 'non-linguistic', McNeill (2001: 3) means that a gesture "is non-morphemic, not realised through a system of phonological form constraints and has no potential for syntactic combination with other gestures". Reformulating McNeill's arguments in SFL terms, gestures are non-compositional (no part-whole constituency), exhibit an arbitrary relationship between content and expression and cannot be syntagmatically organised. Put yet another way, gestures do not have a grammar.

From an experiential perspective, gestures cannot be broken down into Participants, Processes and Circumstances in the way that speech or writing can; that is, there is no 'separate' hand shape for Participant, one for Process and one for Circumstance. Instead, a gesture seems to contain, in Halliday and Matthiessen's (1999) terms, a complete *figure*. McNeill sees gestures as *global*, in that the 'whole' gesture works top-down in determining how its parts may be understood. Furthermore, gestures are in McNeill's view synthetic, in that various meaningful parts such as 'he+bends+it+way+back' are represented as a single gesture. McNeill's claim that gestural form and meaning are arbitrary rests on the observation that various meanings such as 'bending something back' are not consistently realised in specific gesture shapes. So for instance, bending back a tree may be realised in a 'grabbing' gesture that is then moved backwards. By way of contrast, bending back a piece of paper may be realised in a 'pinching' gesture shape that is moved backwards. From this perspective, form and movement can realise a whole host of meanings.

McNeill's final point, that gestures are non-combinatoric, means that gesture shapes do not 'add up' to construct a 'larger' meaning. Instead, gestures are independent of each other, and tied to specific segments of speech. This argument is similar to the non-constituency view of gestures. You cannot, for instance, add meanings of 'manner' or 'location' to a previously performed gesture.

The degree to which gesture content-expression pairs are related is certainly much looser than in speech or writing. As noted above, a certain form is not in any way tied to a specific meaning. But, how 'loose' is this relationship between form and meaning? Is there no constraint whatsoever? To shed more light on this question, we need to examine the range of meanings realised in various forms of expression. Only in this way can we say with some certainty that certain meanings may be ruled out. Showing that a meaning such as 'bending back' can be realised in different forms does not necessarily mean that anything goes. It might mean instead that speakers have a number of options for realising meanings in gesture. The non-combinatoric properties of gestures should also be more critically examined. Although McNeill's research suggests that a sequence of gestures does not add up to construe 'larger' meanings, it may be the case that we do not see these larger meanings, because our categories prevent us from doing so. If we concede to McNeill's claim that gestures are global-synthetic, then we will not want to break up a gesture into component parts; that is, gesture form and movement will not have meanings that are distinct from each other. Global-synthetic means that the 'whole' determines the meaning of the parts. I would suggest that movement and form have meanings that are independent of the whole. And I believe that SFL theory provides a perspective that enables us to construe these meanings.

03 F: Somebody explained to me yesterday how to get the eight ball in on the break,
but . . . it hasn't worked ye@e@et
((Eddie moves to end table, chalking cue, places chalk on table edge))
04 (2.0) ((Eddie places stick over table parallel to side))
Gi. . .xxxxxxxxxxxxxxxxxxxxxxxxxxxx ((Eddie gazes at Franklin))
05 E: Use a top English on the side?
!**********!********!



((left hand, C-shape: thumb/index, twist))

Example 1. Two pool players' recounting of a pool shot

In order to address these issues of constituency, form-meaning pairs, and combinatory potential, I show an example of two pool players' recounting of a pool shot (see Appendix for definitions of the notation conventions used in the transcript; the example is taken from Muntigl 1996: 284). Franklin (F:) and Eddie (E:) have just begun a game of pool. Eight seconds after the break, Franklin initiates a recount of a type of shot made in pool, by suggesting a method of sinking the eight ball on the break as shown in Example 1.

Eddie suggests with rising intonation that the cue ball is given top English (a backspin). At the beginning of the recount, Eddie illustrates what he means by *top English* by producing a gesture to indicate a special spin on the ball. Eddie uses a C-shape gesture with the thumb and index finger of his left hand while twisting his wrist to indicate a ball with a spin.

According to McNeill's (1992) classification of gestures, the C-shape gesture is *iconic*, because its shape represents the pool ball. In SFL terms the C-shape, 'the pool ball', experientially construes a Participant. Furthermore, the twisting motion realises manner, in that it depicts how the ball spins. If we were to follow McNeill, this gesture would be analysed global-synthetically; that is, the gesture would consist of the shape and movement, and, in addition, the meaning of the shape and movement could only be understood in reference to the gesture as a whole. Thus, the twisting motion construes manner only because we know that the hand metaphorically refers to a pool ball. McNeill is certainly correct in asserting that the meaning of the gesture's parts is constrained in a top-down fashion. For instance, if the C-shape was not construed as a pool ball (i.e. if it was not an iconic gesture), then movement would not necessarily have realised manner.

Nonetheless, there is a good argument to be made that the parts also contribute to the meaning of the whole. If the gesture is examined for its experiential meanings, then the C-shape may be construed as the Participant, and the movement or twisting motion as the Process. Furthermore, these experiential meanings may be verbalised ergatively as 'the pool ball + spins + backwards'. In this way, the 'whole' gesture may be interpreted as a *middle* clause (Halliday 1994: 169) in which the gesture form is the Medium and the movement is the ergative Process. Looking at this gesture compositionally, the form is one part (i.e. the Medium), the movement is another (i.e. the Process) and the direction of movement yet another (i.e. Manner).

A C-shape gesture will not, of course, always iconically realise a pool ball or a type of round object. We can all imagine situations in which the same gesture has been used in a spatial sense to construe 'thickness' or 'width', or in a more metaphorical sense in which a speaker is considering two possibilities. In the latter sense, the C-shape might abstractly refer to the two possibilities in question, and the twisting motion might construe the Process of 'contemplating' or 'considering'. In this sense, McNeill is surely right. We have to understand the figure that the gesture realises, before we know what meanings the parts realise. However, in the gesture in (1) I would argue that the parts contribute to our understanding of the meaning of the whole gesture. If, for example, E would not have applied movement to the gesture, it might have been difficult to interpret what the hand is actually signifying. Notice that E does not say, "pool ball". We have to infer this from his gesture. Through the twisting a link is made between "top English"and a spin. Therefore, it is only through the component parts of form and movement that we are able to 'see' a spinning pool ball.

By looking at gestures in terms of SFL categories such as Participant and Process, it becomes possible to read gestures compositionally. Furthermore, it is doubtful that specific gesture forms such as a C-shape can mean just about anything. I think there is a constraint, however minimal, between gestural form and meaning. To conclude, I would suggest that McNeill's proposed properties need to be relaxed. Gesture is a semiotic system that includes a content and expression stratum. Whether or not gestures have a grammar is still open to debate.

2.2 Semiotic systems and metafunctions

The above analysis of the C-shaped gesture has shown how ideational meanings such as Participant, Process and Manner may be mapped onto gestures. The next step would be to see if gestures are also interpersonally and textually organised. The textual metafunction is not difficult to demonstrate. Since gestures occur in a 3D gesture space consisting of six quadrants (see McNeill 1992), gestures may be analysed for where they occur in this gesture space – left, right, top, bottom, front

or back. Some of the textual systems that may be activated are THEME, INFORMATION, and REFERENCE (Halliday 1985) or even IDEAL-REAL (Kress & van Leeuwen 1996). Gestures may also realise interpersonal meanings. For example, McNeill (1992) argues that a certain type of gesture termed *beat* functions to evaluate certain parts of a speaker's message. Other interpersonal functions may include questioning the import of what someone has said (Kendon 2001), regulating speakers' patterns of attention (Goodwin & Goodwin 1986), assessing another's utterance (Goodwin & Goodwin 1992), and projecting the next part of the text (Streeck & Hartge 1992). One important question that arises in gesture research is whether gestures may construe both ideational and interpersonal meanings. If, as in beat gestures, the hand is not functioning as a Participant, then it is questionable as to whether the gesture will, in addition to evaluating the message, realise ideational meanings.

The semiotic system that has probably received the most attention in terms of metafunctional modelling is image. In their book *Reading Images* Kress and van Leeuwen (1996) demonstrate how various ideational, interpersonal and textual meanings are realised in images. Ideational meanings include Participants (e.g. Actor, Goal, Reactor, Phenomenon, Senser, Sayer), Processes (e.g. Mental, Verbal, Action, Reactional) and Circumstances (e.g. Setting, Means, Accompaniment). Interpersonal meanings include contact, social distance, attitude and modality. Finally, textual meanings include the spatial arrangement of information (i.e. Given/New and Ideal/Real). Furthermore, Kress and van Leeuwen argue that images have a grammar. If grammar is associated with metafunction, then they are surely right. In Halliday's (1994) model clause types such as material and verbal, Participants such as Actor and Sayer and Circumstances such as Manner and Accompaniment are *grammatical,* not semantic functions. What is much more difficult to identify in images (and gesture) is a visual/spatial equivalent to a grammatical class such as noun or verb. In order to advance the discussion of how semiotic systems are stratally organised, more precision should be given to what is meant by content (i.e. semantics and/or grammar) and expression. If a semiotic system has a grammar, how can we differentiate between the system's grammar (i.e. how constituent noun-like or verb-like parts are syntagmatically organised) and how the grammar is being expressed? For gestures the components of expression are kinesic and include arm, head and body position and movement, Gesture-Unit and Gesture-Phrase (see McNeill 1992: 82ff.). When these body parts take on functional shapes, it is possible that a grammar becomes realised. One crucial difference between sign language and gesture is that the functional shapes in the former semiotic system – and their possibilities for combination – are much more highly developed than the latter.

3. Mode and the relationships between semiotic systems

Up to now it has been argued that, in addition to language, a variety of semiotic systems such as gesture and image may be deployed to make meaning. What still needs to be addressed is how different semiotic systems or combinations of systems become activated in different social contexts and how combinations of semiotic systems functionally interrelate, by expanding the meaning of the other. Thus, this section begins by arguing that the situational variable *mode* is centrally involved in organising which semiotic systems become activated in different contexts of situation.

3.1 Mode in SFL

In Halliday's framework *mode* is a contextual variable. It is primarily responsible for organising the part that language plays in social action. A slightly modified account of Halliday's (1985) definition of mode is given by Halliday and Martin:

> MODE – the symbolic organisation: 'what role language is playing' refers to what part language is playing, what it is that the participants are expecting the language to do for them in the situation: the symbolic organisation of the text, the status that it has, and its function in the context, including the channel (is it spoken or written or some combination of the two?) ... Halliday (1985:12) (including the degree to which language is part of or constitutive of what is going on and the potential for aural and visual feedback between interlocutors).
>
> (Halliday & Martin 1993:33)

By suggesting that language plays a part in realising social action, it is implied that other semiotic systems might also be involved. Nevertheless, Halliday does not speculate on the meaning-making role of semiotic systems other than language. The above definition may also be productively reworked in terms of Hasan (1985). Hasan identifies three main dimensions of mode: *language role, medium and channel*. *Language role*, for instance, is placed along a continuum between the end-points *constitutive* and *ancillary*. From this perspective the role of language in social action may be seen as the degree to which language is constitutive of the social action. If language plays a minor role, it is ancillary to the social action. For instance, in the service encounter genre certain stages such as Purchase may be non-verbally realised (see also Ventola 1987 for a discussion of the non-verbal realisation of service encounter stages). If, on the other hand, language plays a major role, as is often the case in story genres, it constitutes the social action.

The categories spoken and written are assigned to *medium*. Hasan accounts for these categories, by introducing the notion of process sharing. In the spoken medium interlocutors are able to share in the text creation process. In contrast,

the written medium does not allow for this kind of sharing. Since written texts may be viewed as finished products, the recipients of these texts may not normally partake in text construction. Also important for process sharing is the degree to which immediate feedback is possible. In the spoken medium, since texts are being co-constructed by interlocutors, speakers are constantly receiving verbal and non-verbal feedback from their addressees. Written texts do not allow for the same kind of immediate feedback.

The third aspect of mode, *channel*, refers primarily to visual and aural.[2] Important in multimodal analyses is the relationship between channel and different semiotic systems. Martin (1992), for instance, argues that the availability of channels, visual and aural, will influence the kinds of semiotic systems that may be brought into the meaning-making process. This is an important point to consider in multimodal analyses. For instance, if both visual and aural channels are open, speakers may engage in face-to-face conservation, and may provide each other with both visual and aural feedback. If the aural channel is not available, say in sign language, speakers must primarily rely on visual signs. Martin also points out that only one channel may be available to one of the interactants. Radio, for instance, transmits an aural signal but does not receive one. Therefore, radio relies on one-way aural contact and no visual contact. Silent movies, on the other hand, would rely on one-way visual contact, and no aural contact – unless music is played during the movie. Martin's work highlights the important connection between channel and semiotic systems. In multimodal analyses, consideration should be given to what channels are open and for whom, and which semiotic systems are being deployed.

The concept of channel might also be expanded to include other kinds of modalities such as *smell, touch* and *taste*. In workplace settings the accurate identification of smells, the taste of various foods, or the touch of different materials may be of central importance and could at times play a pivotal role in the construal of experience. Furthermore, the visual channel should include more than just the ability to see one's interlocutor. The availability of the visual channel allows speakers to deploy other meaning-making systems such as *dress, gesture, gaze, colour*, and *body position*. Channel, therefore, should not be too closely tied to speech, but should be sensitive to a full range of semiotic systems.

3.2 Mode as the primary activator of semiotic systems

What I am proposing is that the contextual variable mode plays a pivotal role in organising which semiotic systems become activated in social contexts. In a way this makes sense, if mode is responsible for such categories such as channel, medium and language role. Since mode organises field and tenor choices (Martin

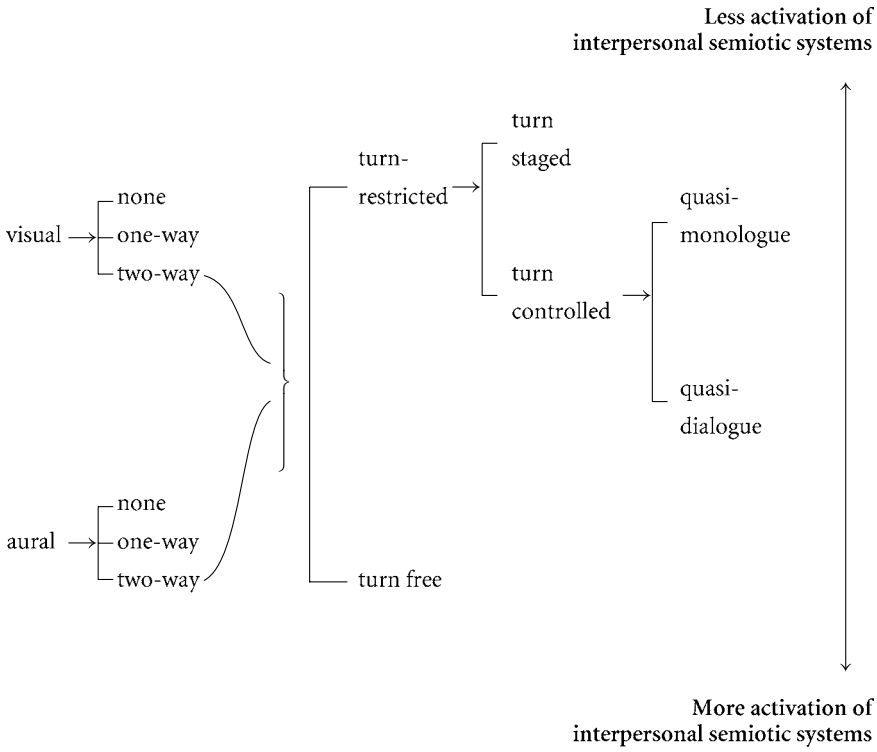


Figure 3. Degree to which turn-taking systems are activated in speech (adapted from Martin 1992: 515)

1992), these contextual variables will also influence the range of semiotic systems deployed. In the following discussion, I will restrict the discussion of mode to face-to-face interaction involving two-way aural and visual communication.

Martin (1992) argues that the degree to which the visual and aural channels are accessible is influenced by the kind of 'medium' used in communicating messages. For instance, a radio operates using only a one-way aural channel. Television and movies, on the other hand, relay messages one-way both aurally and visually. Finally, two-way access of both aural and visual channels is most characteristic of face-to-face conversation. In regard to other semiotic systems such as gesture, eye gaze and body position, it is important to consider how differences in mode may influence the deployment of these systems. Following Martin (1992), some of these differences may be characterised in terms of how tenor or field choices relate to channel. In face-to-face conversation involving two-way aural and visual channels, turn-taking may range along a continuum from free to restricted (see Martin 1992), as shown in Figure 3.

In turn-free interactions meaningful selections from speech, gesture, eye gaze, and body position are constantly being made and negotiated between speakers. However, as the degree of freedom in turn-taking diminishes, the ability to negotiate these semiotic systems diminishes; that is, the way in which these semiotic systems may be used interpersonally, to enact social reality, becomes more restricted. This may become clearer if we take an example of turn-staged interactions such as debates or quasi-dialogues such as speeches or paper presentations. Although the speaker in these contexts uses speech, gesture, body position in a metafunctionally diverse way (i.e. ideationally, interpersonally, and textually), the audience may not. In this way, the speaker's meanings may not be immediately responded to, and hence negotiated; and the interpersonal meanings commonly used to negotiate interaction are at best minimally activated.

The experiential content of the social process may also influence the deployment of semiotic systems. It can be argued that gestures have a more limited experiential scope than speech. Although a large number of gestures may be formed to realise figures of doing, such as material and behavioural clauses, it may be more difficult for gestures to adequately convey different kinds of sensing, saying and being. There is also the question of how well gestures may realise generalised Participants (e.g. classes of things) as opposed to specific Participants (e.g. particular people or objects). Finally, gestures may realise circumstantial meanings such as location and manner and may also realise qualifiers involving quantity, weight and texture.

Other meanings such as *cause*, *extent*, *matter*, *angle* and *role* for circumstances and *colour*, *nationality*, *life* and *age* for qualifiers (see Halliday & Matthiessen 1999: 211 for a classification of qualities) may be more resistant to gestural expression. Based on this tentatively proposed range of ideational meanings for gesture, it is possible to make predictions about which social contexts are most likely to activate these gestural meanings. Martin's (1992) work on *genre* provides one source that enables these kinds of connections. Genres that are realised through activity sequences and/or specific Participants such as stories, procedures, explanations and descriptions are likely contexts for gesturing. Moreover, interactions that realise ideational meanings of location, manner or specific classifications may also activate the gestural system. By way of contrast, genres that are thing-focused, such as many exposition genres, are less likely to activate ideational meanings though gesture. The relationship between activity sequences, participants and the degree to which ideational gestures may become activated is shown in Figure 4.

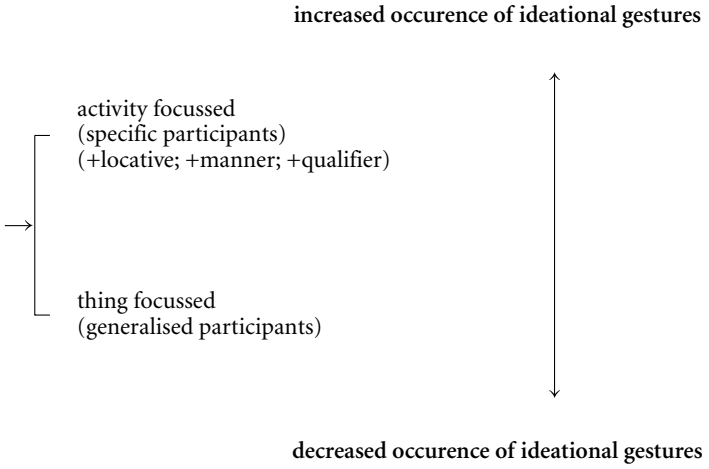


Figure 4. Degree to which ideational gestures are activated in activity-focused vs. thing-focused genres

3.3 Fractal principles between semiotic systems

So far it has been argued that, following Halliday and Matthiessen (1999), semiotic systems other than speech or writing are stratally and metafunctionally modelled in the image of language; that is, semiotic systems such as gesture may in different ways 'add' textual, interpersonal, and ideational meanings to speech. One way of modelling the functional interrelationships between semiotic systems is to follow Halliday and Matthiessen's (1999) and Martin's (1995) suggestion in proposing a set of general motifs that run throughout the grammar's construal of experience. What they are claiming is that expansion and projection are *transphenomenal* categories or *fractal* types, in that they pervade the whole semantic space of the ideation base; that is, *fractal* types of expansion and projection operate not only within *sequence*, but also within the semantic space of *figure* (i.e. clause) and *element*. These fractal types are also said to operate between genres and the stages within a genre (Martin 1995). In this way fractal types operate at different levels within a semiotic system (e.g. language), or from a multimodal perspective, between different semiotic systems, such as speech and gesture. Ideationally, fractal types of projection include locution (") and idea ('), whereas categories of expansion include elaboration (=), extension (+) and enhancement (x). Categories of projection and expansion are shown systemically in Figure 5.

A brief summary of each logico-semantic type is given in Figure 6, taken from Halliday (1994: 220).

By putting together the idea of fractal types, such as projection and expansion, operating between semiotic systems and the idea of social contexts activating

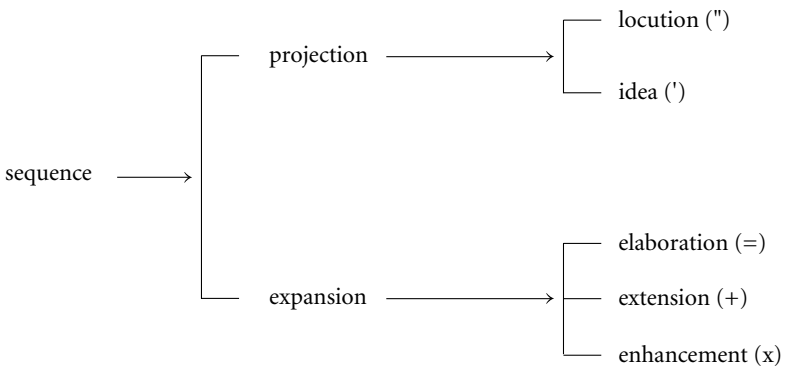


Figure 5. A system network for sequence

Elaboration *i.e.*	one clause expands another by elaborating on it (or some portion of it): restating in other words, specifying in greater detail, commenting, or exemplifying.
Extension *and, or*	one clause expands another by extending beyond it: adding some new element, giving and exception to it, offering an alternative.
Enhancement *so, yet, then*	one clause expands another by embellishing around it: qualifying it with some circumstantial feature of time, place, cause or condition.
Locution *says*	one clause is projected through another, which presents it as a locution, a construction of wording.
Idea *thinks*	one clause is projected through another, which presents it as an idea, a construction of meaning.

Figure 6. Brief summary of logico-semantic types after Halliday (1994: 220)

combinations of semiotic systems, we may come up with the model represented in Figure 7.

The model shown here only reveals the transphenomenal categories of the ideation base operating between gesture, speech and eye gaze. A more complete model would, of course, include interpersonal (i.e. prosodic) and textual (i.e. periodic) transphenomenal categories. Following Martin (1992) and Ventola (1987), genre is treated as a separate socio-semiotic system that ranges above field, tenor and mode. Genre activates patterns of field, tenor and mode variables. And, taken together, social context will activate patterns of meanings that are realised in combinations of semiotic systems. Although I have argued that mode – in terms of channel, language role and medium – is primarily responsible for activating these patterns, field, tenor and certainly genre will influence which semiotic systems become activated. Furthermore, semiotic systems combine their meanings in terms

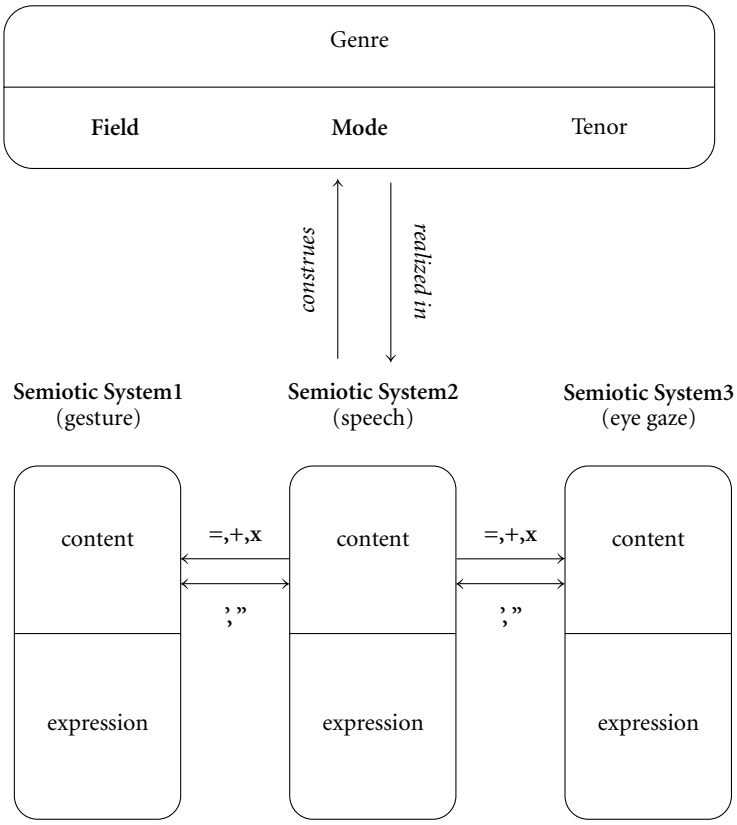


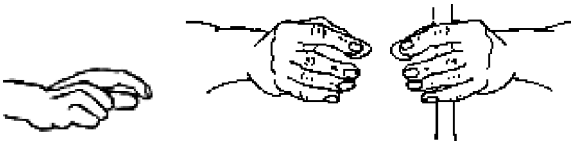
Figure 7. The relationship between social context and semiotic systems

of transphenomenal categories. Ideationally, these transphenomenal categories include projection (', ") and expansion (=, +, x).

4. Conclusion: An example using gesture

To conclude this chapter, I show how the model represented in Figure 7 may operate for speech and gesture. In order to do this, I return to the C-shaped gesture of Example 1. Recall that, in this example, the speakers were engaged in a recount genre. According to Martin (1992), recounts involve a Record stage in which a series of activities involving specific participants are expressed. For these reasons one might expect gesturing to occur. Furthermore, the field involves the game of pool and, more specifically, pool shots. Again, the field would predict the realisation of activities, specific Participants (e.g. pool ball, table, rack, cue, etc.) and meanings

Gixxxxxxxx.....xxxxx
18 E: it'll hit the– hit the rack,
!********!*****_____!



away rack

((extends arm and jerks C-shapes on "hit", opens hands, brings hands together in triangular formation on "rack"and holds))

Example 2. Recount of how cue ball proceeds towards the rack with a top English spin

of Manner and Location. Indeed, as Example 1 shows, the gesture shape construes Manner and a Participant. In reference to channel, language role and degree of turn-taking, we might also predict the production of gestures. The speakers are co-present, have access to both the aural and visual channels, are in a turn-free interaction, and are using semiotic systems to constitute the social activity that they are in. But what of the relationship between what E says and what he gestures? I would argue that the relationship is of the expansion: enhancement (x) type, in which the twisting C-shaped gesture depicts *the result* of using a top English. A paraphrase may be worded as "use a top English on the side, $^{(x)}$*so that* the cue ball will spin."In this way, the gesture is able to expand the meanings of speech through manner (twisting C-shape = spinning ball) and cause (the gesture is the result of using top English).

Another type of functional relationship is shown in Example 2 – taken from Muntigl (1996: 285) – appearing somewhat later in the recounting of the pool shot. While E recounts how the cue ball proceeds towards the rack with a top English spin (*it'll hit the– hit the rack*), he also gestures by making a triangular shape with his hands.

In Example 2, the gesture expands the meanings of speech through elaboration (=). More specifically, the triangular-shaped gesture elaborates on the meaning of *the rack* by describing its shape. A paraphrase might be: "it'll hit the– hit the rack, $^{(=)}$which at this point has a triangular shape."

In summary, multimodal analyses need to account for the role of mode in meaning-making – how many channels, what is the role of each semiotic system, etc. – and how combinations of semiotic systems functionally interrelate in different contexts of situation. What this means is that we should broaden our analytic lens beyond speech and writing to include other meaning-making systems.

Notes

* I gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada, Postdoctoral Fellowship No. 756-2001-0224 in the preparation of this manuscript.

1. For a notable exception, see Huemer (2001), who examined the functional interrelationships between images and writing in job advertisements.

2. Hasan's (1985: 58) definition of channel "refers to the modality through which one comes in contact with the message – whether the message travels on sound waves or on a piece of paper". Hasan argues for two types of channels through which messages may travel: *phonic* or *graphic*. Unfortunately, the term graphic privileges *writing* as a semiotic system, and leaves out gesture, image, body position and eye gaze as meaningful contributors to the meaning-making process. For this reason, I adopt visual-aural rather than phonic-graphic.

References

Goldin-Meadow, Susan & Mylander, Carolyn (1984). "Gestural communication in deaf children: the effects and non effects of parental input on early language development." *Monographs of the Society for Research in Child Development, 49* (3), 207.

Goodwin, Charles (1994). "Professional vision." *American Anthropologist, 96* (3), 606–633.

Goodwin, Charles & Goodwin, Marjorie H. (1992). "Context, activity and participation." In P. Auer & A. di Luzio (Eds.), *The Contextualization of Language* (pp. 77–99). Amsterdam: Benjamins.

Goodwin, Charles & Goodwin, Marjorie H. (1996). "Seeing as situated activity: formulating planes." In Y. Engeström & D. Middleton (Eds.), *Cognition and Communication at Work* (pp. 61–95). Cambridge: Cambridge University Press.

Goodwin, Charles & Goodwin, Marjorie H. (1986). "Gesture and co-participation in the activity of searching for a word." *Semiotica, 62*, 51–75.

Halliday, M. A. K. (1978). *Language as a Social Semiotic.* London: Edward Arnold.

Halliday, M. A. K. (1992). "How do you mean?" In M. Davies & L. Ravelli (Eds.), *Advances in Systemic Linguistics* (pp. 20–35). London: Pinter.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M. A. K. (1985). "Context of situation." In M. A. K. Halliday & R. Hasan (Eds.), *Language, Context and Text: Aspects of Language in a Social Semiotic Perspective* (pp. 3–14). Geelong, Victoria: Deakin University Press.

Halliday, M. A. K. & Martin, James R. (1993). *Writing Science: Literacy and Discursive Power.* Pittsburgh: University of Pittsburgh Press.

Halliday, M. A. K. & Matthiessen, Christian M. I. M. (1999). *Construing Experience through Meaning: A Language-Based Approach to Cognition.* London: Cassell.

Hasan, Ruqaiya (1985). "The structure of a text." In M. A. K. Halliday & R. Hasan (Eds.), *Language, Context and Text: Aspects of Language in a Social Semiotic Perspective* (pp. 52–69). Geelong, Victoria: Deakin University Press.

Huemer, Birgit (2001). Aufbau und Funktion von Stellenanzeigen: Eine funktional semiotische Analyse. Unpublished M.A. Thesis. University of Vienna.

Hutchins, Edwin & Klausen, Tove (1996). "Distributed cognition in an airline cockpit." In Y. Engeström & D. Middleton (Eds.), *Cognition and Communication at Work* (pp. 15–34). Cambridge: Cambridge University Press.

Kendon, Adam (1982). "The study of gesture: Some remarks on its history." *Recherches Sémiotique / Semiotic Inquiry, 2*, 45–62.

Kendon, Adam (1985). "Behavioral foundations for the process to frame attunement in face-to-face interaction." In G. P. Ginsberg, M. Brenner, & M. von Cranach (Eds.), *Discovery Strategies in the Psychology of Action* (pp. 229–253). London: Academic Press.

Kendon, Adam (1988). "How gestures can become like words." In F. Poyatos (Ed.), *Cross-cultural Perspectives in Non-verbal Communication* (pp. 131–141). Toronto: Hogrefe.

Kendon, Adam (1990). *Conducting Interaction: Patterns of Behavior in Focused Encounters.* Cambridge: Cambridge University Press.

Kendon, Adam (1997). "Gesture." *Annual Review of Anthropology, 26*, 109–128.

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images.* London: Routledge.

Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Edward Arnold.

Lebaron, Curtis & Streeck, Jürgen (1997). "Built space and the interactional framing of experience during a murder interrogation." *Human Studies, 20*, 1–25.

Martin, James R. (1992). *English Text*. Amsterdam: John Benjamins.

Martin James R. (1995). "Text and clause: Fractal resonance." *Text, 15* (1), 5–42.

Matthiessen, Christian M. I. M. (1993). "Register in the round: diversity in a unified theory of register analysis." In M. Ghadessy (Ed.), *Register Analysis: Theory and Practice* (pp. 221–292). London: Pinter.

McNeill, David (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, David (2001). "Introduction." In David McNeill (Ed.), *Language and Gesture* (pp. 1–10). Cambridge: Cambridge University Press.

McNeil, David, Cassell, Justine, & McCullough, Karl-Erik (1994). "Communicative effects of speech-mismatched gestures." *Research on Language and Social Interaction, 27* (3), 223–238.

Muntigl, Peter (1996). "An analysis of the interactive organisation of shot discussions during pool games." In *Proceedings of the 1996 Annual Conference of the Canadian Linguistics Association*, *Calgary Working Papers in Linguistics* (pp. 281–291).

Streeck, Jürgen (1993). "Gesture as communication I: Its coordination with gaze and speech." *Communication Monographs, 60* (4), 275–299.

Streeck, Jürgen (1994). "Gesture as communication II: The audience as co-author." *Research on Language and Social Interaction, 27* (3), 239–267.

Streeck, Jürgen & Hartge, Ulrike (1992). "Previews: Gestures at the transition place." In P. Auer & A. di Luzio (Eds.), *The Contextualization of Language* (pp. 135–157). Amsterdam: Benjamins.

van Leeuwen, Theo (1999). *Speech, Music, Sound.* London: Macmillan.

Ventola, Eija (1987). *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters.* London: Pinter.

Appendix

Notation conventions

(8.0)	length of pause
((breaks))	comments
(okay)	transcriptionist doubt
Gi	initiation of gaze
Gf	end of gaze
. . . . .	gaze towards and away
xxxxxx	gaze at addressee (Eddie's gaze at Franklin)
!	beginning or end of gesture
******	stroke of gesture
_____	holding gesture
[	overlap
:::	lengthening
@@@	laughter

Chapter 3

Problematising 'semiotic resource'*

Victor Lim Fei
National University of Singapore

This chapter investigates the nature of a semiotic resource and systems from the Systemic Functional Linguistics (SFL) perspective. A semiotic resource contains both an expression and content plane as well as possesses system networks on each of these planes. A system, on the other hand, is a configuration of meaning potentials that is articulated through a semiotic resource. This is followed by an argument of a visual image as a semiotic resource, comparing it with the modality of language. In addition, it will also be proposed that Saussure's (1967) claim of arbitrariness between a signifier and a signified could further be extended in the current understanding of the nature of language and visual images, especially since both semiotic resources share a common historical origin. The implications arising from the association between the two modalities are also discussed briefly. Stemming from a need to understand the semiotic resource of visual images, this chapter also proposes icons as the vocabulary of visual images, analogous to the role of words in language. The conceptions presented in this chapter are preliminary and by no means final nor definitive. The chief aim of this chapter is to provoke a meaningful debate on some of the pertinent questions in multimodal research.

1. Multimodal semiotic resources

Earlier work on meaning has centred on the notion of the sign. This focus is only shifted when the work of Michael Halliday on Systemic Functional Linguistics theory (SFL), redefines the boundaries of semiotics from "a study of signs" to "a study of sign systems". Halliday's (1978) work marks a shift of emphasis from a sign as an entity, to a system of signs operating together to make meaning. In the SFL community signs are more commonly referred to as semiotic resources. They include language, expressed in its written form through graphology or typography, as well as the semiotic resources of visual images, mathematical notations and other technical symbols.

In this age of the multimedia, there is an increasing awareness that meaning is rarely made with language alone. As Baldry (2000) and Kress and van

Leeuwen (2001) note, we live in a multimodal society, which makes meaning through the co-deployment of a combination of semiotic resources. Visuals, gestures and sounds often accompany the linguistic semiotic resource in semiosis. As such, there is a pressing need to understand the dynamics of meaning-making, or semiosis, in multimodal discourse. Academic disciplines that focus on monomodality, such as linguistics, must come into dialogue with other fields of research, for instance, visual communication studies and media studies, to facilitate the interdisciplinary nature of multimodal research.

Kress expresses the pressing need to understand the role of the visual semiotic in communication, arguing that "it is no longer possible to avoid these issues in critical analyses, on the assumption, explicitly or implicitly held, that all (relevant) meaning in a text is, as it were, fully glossed in the verbal component of the text" (Kress 1993: 188). Research in multimodal discourse analysis has also been increasing (see, Baldry 2000; Kress & van Leeuwen 2001; O'Halloran 2004). Most of these studies focus on the dynamics and the meanings that result from the co-deployment of the various modalities. Less research efforts have, however, been spent on examining some of the assumptions underpinning multimodal studies, many of which borrowed largely from the extensively documented semiotic resource of language. For instance, the theoretical dimensions of what constitutes a sign and the nature of a semiotic resource need to be re-examined in the light of the more recent advances made in multimodal research. By discussing some of these questions, this chapter shall provoke a deeper exploration into some of these issues in the new field of multimodality.

2. Nature of a semiotic resource

Let us first begin by initiating a discussion on the notion of a visual message, in particular, discourse that co-deploys the semiotic resources of language and visual images, from the Systemic Functional Linguistics (SFL) perspective. Next, it will also be proposed that visual images can be seen as a legitimate semiotic resource, analogous to language. In order to recognise that, it is helpful to clarify the understanding of what is a semiotic resource, as used in this chapter, amongst the many flagrant proliferations of the term.

Apart from pioneering a fresh approach to the study of meaning through the investigating of the relationship between sign systems or semiotic resources, Halliday (1978) also foregrounds the notion of choice, i.e. the selection of an option over others in meaning-making. Halliday (1978) operates on the assumption that a semiotic resource must have (1) an expression and content plane, as well as (2) possess systems operating on each plane. Within each plane, there is a network

of options as "a representation of the potential at that level" (Halliday 1978: 40). A system, according to Halliday, is "a set of options together with an entry condition, such that if the entry condition is satisfied, one option from the set must be selected" (Halliday 1969: 253; quoted in de Joia & Stenton 1980: 109). In other words, the system is "an abstract representation of paradigm" (Halliday 1971: 55; quoted in de Joia & Stenton 1980: 109). The notion of language as semiotic resource is derived from Halliday's (1978) seminal proposition of language as a social semiotic. A social semiotic is "a system of meanings that constitutes the 'reality' of the culture" (Halliday 1978: 123). Semiotic resources are therefore inextricably linked to context and a social reality.

Following Halliday (1978), a guideline on the nature of semiotic resources and systems from the SFL perspective can be established. A semiotic resource possesses a content plane, where a set of grammar operates, and an expression plane, where the content plane is articulated. In addition to that, systems operating within a network are also present on each of the strata or planes. A semiotic resource is thus differentiated from a system, in that a system does not possess a content and expression plane. Instead, systems are metafunctionally based configurations of the meaning potential of each semiotic resource. Systems may also host several *subsidiary systems*, or *sub-systems*, where the sub-systems operate on a different level of delicacy from the main system. For instance, language as a semiotic resource has on its expression plane the system of *typography*. Within *typography*, there is a sub-system of *font*, and within *font*, there are many sub-systems as well, amongst them, for instance, *colour*. With these principles as guidelines, it is easier to distinguish between a semiotic resource and a system or other mechanisms that may all contribute, albeit in different capacity, to the meanings in a text.

A simple example of a semiotic resource, as illustrated by Eggins (1994) is the traffic light. It is a modality, as it possesses both an expression plane and a content plane. The expression plane consists of the system of *colour*, namely, Red, Amber and Green. There is also a set of grammar in operation. This is evident through the fixed sequence in the lighting of colours, and that at any one time only one colour is lit, in effect, a paradigmatic selection is made. The system of *colour* is meaningful but, in itself, it is not a semiotic resource, but rather the system by which the representational meaning potential of the semiotic resource of the traffic light discourse is realised. Other semiotic resources include sculptures (O'Toole 1994), music (Callaghan & Mcdonald 2002), mathematics (O'Halloran 2000) and visual images (O'Toole 1994; Kress & van Leeuwen 1996). Of particular interest in this chapter is the semiotic resource of visual images. A discussion based on a comparison of language and visual images as modalities is initiated in Section 3.

Keeping with Halliday's application of the term 'semiotic resource', it is necessary therefore to be more judicious in the use of this term in multimodal re-

search so as to minimise terminological confusion. Royce, for example, proposes an 'intersemiotic complementarity' that describes the operations of 'intersemiotic resources' (Royce 1998: 45) to produce in the reader, the feel of a single, coherent multimodal page. Unhelpful implications may arise in the naming of the processes responsible for the synergistic combination between the two modalities as 'intersemiotic resources', suggesting that there is a set of semiotic resources responsible for intersemiotic relations. This will lead to the question of what then are the expression and content planes of these 'intersemiotic resources' as well as what are the systems operating within these resources.

I propose that the nature of the processes at work in the co-deployment of modalities may not adhere to the characteristics of semiotic resources, therefore labelling these processes as intersemiotic resources may be counter productive. Following O'Halloran (1999), I have found it more productive to refer to processes such as 'semiotic metaphors' (O'Halloran 2003) and 'homospatiality' (Lim 2004) simply as *mechanisms.* They are responsible for the multiplication of meaning (Lemke 1998) that arises in the synergistic co-deployment of various semiotic resources. The more neutral naming of these processes as mechanisms, frees them from the complications of the association as semiotic resources, as well as serves to establish a theoretical platform where further unhindered investigations into the nature of these processes can take place.

Likewise, van Leeuwen (2002) also notes that typography has been increasingly seen as a semiotic mode in its own right, although van Leeuwen observes that typography does not appear to be quite a stable semiotic system yet. Van Leeuwen's reservation is reasonable. I foresee that typography may never arrive to the stability as a semiotic resource, for instance, like language, because of the nature and disposition of typography. In my approach to the understanding of a multimodal text (Lim 2004), I have found it more helpful to view typography, not as a semiotic mode, but as a configuration of systems, with many sub-systems operating within it. *Typography* is thus a system network on the expression plane of the semiotic resource of language. Similarly, it is more productive to classify *lines, shading* and *shape* not as semiotic resources but as systems and sub-systems through which the pictorial modality expresses meaning. These systems represent a configuration of meaning potential of semiotic resources, having within them paradigmatic options, instead of an expression and a content plane.

3. Visual images as semiotic resources

Having discussed the nature and characteristics of semiotic resources, it is appropriate to examine if the claim of visual images as semiotic resources is tenable.

Comparing the visual images, with the semiotic resource of language, the visual images can be observed to have an expression plane (display stratum) and a content plane (the grammar and semantics strata). Halliday (1978: 39) proposes that language is a "system of meaning potential". Operating on the content and expression plane, each plane has a network of options where meaning is made through the paradigmatic selections. Language is an abstraction until it is expressed through either speech or writing. When the linguistic semiotic is expressed through sound, the display stratum is phonology. When language is realised through writing, the expression plane is graphology or in the instance of a printed text, typography.

O'Toole (1994) and Kress and van Leeuwen (1996) argue that visual images are tools or semiotic resources, just as competent as language, in meaning-making. The adoption of the stance that both the linguistic and pictorial modalities should share an equal status is now widely recognised (for example, Baldry 2000; O'Halloran 2000; Thibault 2000; Kress & van Leeuwen 2001). Van Leeuwen (2000), for instance, criticises the negative comparisons between language and visual images in his refutation of Barthes' (1977) earlier proposition that words have 'fixed meaning' while images are 'polysemous'. In addition to this, van Leeuwen (2000) confronts some misconceptions regarding the pictorial semiotic such as the assertion that visual images cannot represent negative polarity. Van Leeuwen (2000: 179) also argues that visual semiotics should focus, "not only on the image as representation, but also on the image as (inter)act".

I add to these conceptions by proposing that visual images, like language, are conceptual abstractions, each with its potential of meaning. As shown in Figure 1, language is an abstract system of meaning potential, realised through its grammar, and this is expressed on the display stratum, through typography in printed texts. In the same manner, visual images are also abstractions that are realised through a visual grammar network. On the display stratum they are expressed through visual systems of *graphics*, such as *form, perspective, layout* and *strokes*.

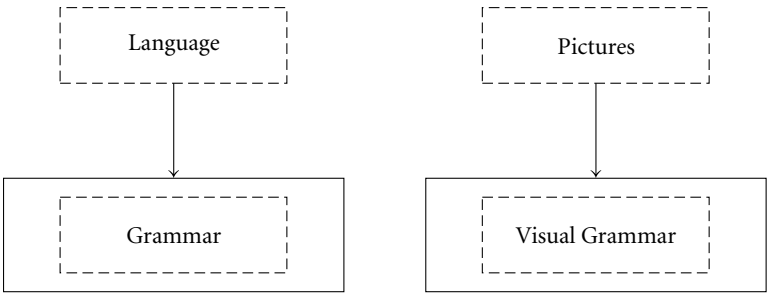


Figure 1. Instantiation of language and pictures (reproduced from Lim 2004)

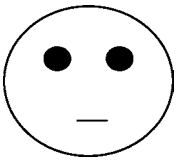


Figure 2. An iconic face (reproduced from Lim 2004)

The separation between display and grammar for the pictorial semiotic may be an uneasy one, due to the interwoven nature of the elements on both strata in semiosis. Nonetheless, it is useful and necessary to distinguish between the two planes, to recognise the systems' potential as well as to understand the meaning-making process. The example in Figure 2 demonstrates the theoretical distinction between the display and grammar strata.

The expression plane of the iconic face in Figure 2 involves the system of *colour* and *form* used to make meaning. This refers to the thin black line, the two black circles as well as the larger white circle. Each of these elements independently as well as together as a unified whole, has meaning potential. The grammar stratum, as extensively theorised by O'Toole (1994) and Kress and van Leeuwen (1996), relates one disparate element to another and explains how the whole functions cohesively to make meaning. Just as the grammar of language concerns itself with the chains of words to form coherent sentences, the grammar of visual images is about the piecing of one item with another to bring across a coherent message. The relation of the parts into wholes, for instance, how the various shapes form an iconic face, operates on the grammar stratum. This grammar is culturally dependant and governs the way a reader 'reads' and construes the visual message.

Following O'Toole (1994), a hierarchy of different ranks analogous to Halliday's (1978) rank scale for language is proposed to look at the meaning made on each of the rank units, from Member to Figure and Episode to Work. This adoption of a rank scale, operating within the principle of constituency, where wholes on each rank make up larger units in a hierarchy, facilitates a more systematic analysis of the meaning made in the different units on the visual grammar stratum. In a sense, this delicate distinction between the display and grammar stratum can be made, with the expression plane being largely concerned with the surface features of the text and the content plane having to do with the interactions and negotiations between the different elements in the text.

4. Extending the claim of arbitrariness

Having established visual images as semiotic resources, analogous to language, it could be useful to look at some of the similarities and differences in nature between the two of them. Eisner (1990: 13) cites evidence claiming that words and visual images "are derivatives of the same origin". It is hardly surprising then that words and visual images share many similarities between them, being derived from a common ancestry. Diringer (1986) further elaborates that pictographs and pictograms are essentially the pre-embryonic stage of writing. The final stage, according to Senner, is achieved

> when pictograms, logograms, and ideograms became phonograms, that is, when the phonetic value of the sign became independent of the original referent of the sign (and eventually of the external shape) and could be combined productively in a conventional system to intercommunicate. (Senner 1989: 5)

Despite sharing the same historical origin, and many similarities, I argue that the difference between the language and visual images lies in the degree of arbitrariness in the relationship between the signifier, particularly the expression plane of the semiotic resource, and the signified, the concept that is represented. Saussure (1967) proposes that there is an *arbitrary relationship* between the signifier and the signified in language. In contrast, Kress (1993: 173) tends towards the other extreme arguing that "the relations of signifier to signified, in all human semiotic systems, is always motivated, and is never arbitrary". He also suggests that production factors such as the 'interest' of the producers, which is subjected to temporality, society and culture, plays a crucial role in the organisation of the sign. More recently, Kress and van Leeuwen (2001) also take into consideration the strata of design, production and distribution in their discussion of multimodal communication.

I propose that the claim of arbitrariness between the signifier and the signified can be further extended. Leaning more towards the interpretation rather than the articulation of the sign in this chapter, I place a greater emphasis on the meanings that can be obtained through the reading of a sign, as opposed to the meanings the sign was produced to convey. Sardar and van Loon (2000: 44) define reading (in the field of media studies) as "the process of interaction when a text is analysed as well as the final result of that process, the interpretation." My stance is inclined towards the post-structuralist position that meaning is found within the unregulated play of *reading* the text, through the interpreting of various semiotic systems, as elaborated in the works of Roland Barthes (1977), Umberto Eco (1979) and Jacques Derrida (1976). The post-modernistic position also paves out the understanding that a text *means* independently of authorial intentions and could be analysed as an artefact of culture.

Taking this position therefore, I hesitate to commit to Kress' (1993:173) claim that all signs are "never arbitrary". His argument that signs are 'motivated' from the perspective of the producer's interest can also be problematised. A question that may be raised from Kress' (1993) discussion is that if the sign is so "opaque" or so inaccessible to the reader to the extent that the 'motivated' meanings are missed, can the sign still be considered as a legitimate sign, now that it fails in its function to effectively communicate meaning. This is seen in his example of the child's drawing of a car (Kress 1993:172). As Kress (1993:178) comments rather paradoxically, "without that accidental presence [of an adult with the child producer] neither interest nor motivation would be easily recoverable". In other words, the intended 'motivated' meaning of the sign may never have been understood by a reader. In such cases then, can the drawing still be considered as a sign when there is no reader that can access the meaning of the sign? Peirce (1958:228) regards a sign as "something which stands to somebody for something in some respect or capacity."

For the purpose of this chapter, following Peirce (1958), I interpret the sign as a tool that facilitates communication, thus necessitating both producer and reader to share the same assumptions and thereby understanding the meanings made through the shared semiotic modalities within a community. Hence, I prefer to view the relations between the sign and the object it signifies on a scale of arbitrariness with the conception of *codification* on one end and the notion of an *analogon* on the other end of the scale.

The signifier of language could be expressed either through sounds in phonemes, in the spoken form, or visually through typography or graphology, expressed in the written form. Concerning the spoken form, that is speech, it is irrefutable that the relation between the signifier and the signified is arbitrary. This is stated with the exception of onomatopoeia (sound words), where the signifier mimics the vocalization of the signified, for instance, the 'ringing' of a telephone. The claim of arbitrariness is also valid in writing systems of language, where the signifier belongs to the syllabic and alphabetic type. The concept of a female child, for example, could be realised by different signifiers in different languages. For instance, in English the signifier is "Girl", in French it is "Fille", and in Italian it is "Ragazza". The lack of an obvious physical relationship between the signifier 'girl' and the signified concept of a 'female child' indicates that their connection is capricious. However, in certain writing systems such as the logographic type, where the signifiers are derived from icons of the objects represented, this claim of arbitrariness may perhaps need to be modified. Certain types of writing systems for language, although having each symbol representing a morpheme, may have their signifiers originating from pictograms, evolving into a standardized writing system over time. Some prominent examples are Chinese Characters and Egyptian

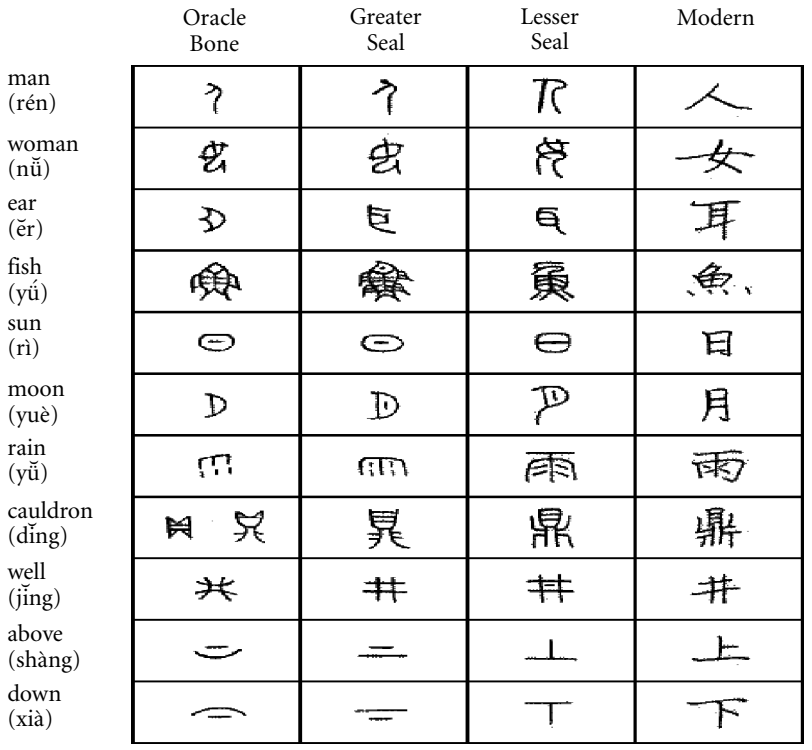


Figure 3. Evolution of 11 Chinese characters through time (reproduced from Keightley 1989)

Hieroglyphics. Tracing the history of such logographic writing systems can illustrate the standardisation and codification of pictograms into a writing system of language over time. Some instances of this are seen in Figure 3.

With this, it is perhaps appropriate to propose differing degrees of arbitrariness between the signifiers and the signifieds in language. As opposed to language, visual images have a lower degree of arbitrariness, thus implying a higher degree of iconicity. Visual images, however, are primarily iconic; that is, they resemble the subjects they represent. Barthes (1977) proposes the term the *perfect analogon* to describe the highest possible level of iconicity or mimesis with the object, such as the image that a photograph produces. In visual images, where there is a higher level of iconicity, the signified and the signifier are related through *mimesis* or resemblance. The opposite end of the scale as opposed to the analogon is the *abstraction*. The analogon has a lower degree of arbitrariness whereas the abstraction has a higher degree of arbitrariness. The typography/graphology of a language is usually the abstraction. Scientific and mathematical notations are those that also lean towards the higher ends of abstraction. Expressionist paintings such as works

of Picasso, for instance, will fall about midway between the scale of abstraction and iconicity.

Since abstractions are characterised through a lower level of iconicity but a higher degree of arbitrariness, the relationship between the signifier and the signified is reinforced through *codification*. In other words, codification links the signifier with the signified sharing a high degree of arbitrariness between them. Codification or 'grammaticalisation' can only take place though effective socialisation into the semiotic community. The term *semiotic community*, follows from Labov's (1972) 'speech community', and describes the people in the same culture, sharing the same assumptions, and selecting choices within the common semiotic resources to make meaning. For instance, in mathematical notations, there is a higher degree of arbitrariness between the signifier and the signified and therefore, stronger codification is required, thus necessitating a deeper initiation of members into the particular community. Notations such as π and Σ can be baffling for the non-members, and their dense meanings are only accessible to members of the particular semiotic community.

5. Basic building blocks: Words and icons

Just as the building blocks of meaning in language are lexical items or words, I propose that the building blocks of visual images are *icons*. In addition, the lexicogrammar of a certain language is culturally specific. For instance, a speaker of Chinese deploys a different lexicon than a speaker of English. Likewise, icons are contextually and culturally specific as well. Different semiotic communities would have different styles of representing the same objects and ideas.

However, the question of where to delineate the boundaries of an icon may arise. For instance, with reference to Figure 2, when is a dot recognised as merely a dot, and when is it functioning as an iconic eye? Icons are the pictorial representations of objects identifiable in the culture. Thus, the recognition of an icon as resembling an item is crucial in deciding what constitutes an icon. The arrangement of lines and dots in a certain manner or 'visual-grammatical' placement, for instance in Figure 2, may bring about the recognition of an iconic face. This identification of the icon is dependant on the relationship between its surrounding *co-text*, in this example, the lines and the dots. The identification of the icon, in turn, also allows us to recognise the iconicity of these co-texts. For example, after recognising the iconic face, the significance of its co-text becomes apparent, for instance, it is clear that the dots represent the eyes and the line stands for the mouth. This is similar to how certain ambiguous words in language are disambiguated when construed in relation to their surrounding co-text, i.e, the other words sur-

rounding them. For instance, the word "bank" can either mean the sides of a river or a financial institution. When used in "The robbers broke into the bank", the polysemeousness of the word is disambiguated. It must be clarified that the internal arrangements of lines and dots to constitute an icon are part of its visual grammar, just as the relationship between part and whole is the grammar of the semiotic resource. In other words, the icon itself lies on the expression plane of the modality, although the composition of an icon and the relationship between iconic elements belong to the grammar stratum.

Due to constraints of time and space, it is not possible for a detailed investigation into the different implications of the proposal of icons as the vocabulary of visual images to be undertaken here. Nonetheless, my proposal will hopefully initiate further work along this direction, which can contribute to a better understanding of the nature of visual images as semiotic resources.

6. Conclusion

This chapter has provided an ambitious attempt to address the complexities of some of the issues raised by multimodal research. The recognition of multimodality as a significant aspect of meaning-making in text has ushered many issues and questions pertaining to multimodal research. This chapter has been titled as *Problematising 'Semiotic Resource'*, because its main objective has been to raise some questions, many of which are concerned with the nature of a semiotic resource, and to examine some of the implicit assumptions in multimodal research. A further aim of the chapter has been to initiate an interest in the semiotic resource of visual images, especially the systems and processes that operate on its expression and content plane. Although some attempts at answers have been proposed in this chapter alongside with the questions raised, the answers presented here are rudimentary and far from final. The purpose of this chapter has not been to offer any simple solutions, but rather to highlight some relevant issues concerning the visual message and the multimodal text, as well as to stimulate a meaningful discussion stemming from the ideas proposed here.

Note

* I would like to thank Kay O'Halloran for her insightful comments on an earlier draft of this chapter.

References

Barthes, Ronald (1977). "The rhetoric of the image." In S. Heath (Ed.), *Image, Music, Text* (pp. 32–51). London: Fontana Press.

Baldry, Anthony P. (2000). *Multimodality and Multimediality in the Distance Learning Age.* Campobasso: Palladino Editore.

Callaghan, Jean & McDonald, Edward (2002). "Expression, content and meaning in language and music: An integrated semiotic analysis." In P. Mckevitt, S. Ó'Nualláin, & C. Mulvihill (Eds.), *Language, Vision and Music* (pp. 221–230). Amsterdam and Philadelphia: John Benjamins.

Derrida, Jacques (1976). *Of Grammatology.* Baltimore: Johns Hopkins University Press.

Diringer, David (1986). *The Alphabet: A Key to the History of Mankind*. New York: Funk & Wagnalls.

Eco, Umberto (1979). *The Role of the Reader: Explorations in the Semiotics of Texts.* Bloomington: Indiana University Press.

Eggins, Suzanne (1994). *An Introduction to Systemic Functional Linguistics*. London and New York: Pinter Publishers.

Eisner, Will (1990). *Comics and Sequential Art*. Princeton, WI: Kitchen Sink Press.

Halliday, M. A. K. (1978). *Language as A Social Semiotic*. London: Edward Arnold.

de Joia, Alex & Stenton, Adrian (1980). *Terms in Systemic Linguistics: A Guide to Halliday*. London: Batsford Academic and Educational Ltd.

Keightley, David N. (1989). "The origins of writing in china: scripts and cultural contexts." In W. M. Senner (Ed.), *The Origins of Writings* (pp. 160–170). Lincoln, NE: University of Nebraska Press.

Kress, Guenter (1993). "Against arbitrariness: the social production of the sign as a foundational issue in critical discourse analysis." *Discourse and Society, 4* (2), 169–191.

Kress, Guenter & van Leeuwen, Theo (1996). *Reading Images: The Grammar Of Visual Design.* London: Routledge.

Kress, Guenter & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication.* London: Arnold.

Labov, William (1972). *Sociolinguistics Patterns*. Philadelphia: University of Pennsylvania Press.

Lemke, Jay L. (1998). "Multiplying meaning: visual and verbal semiotics in scientific text." In J. R. Martin & R. Veel (Eds.), *Reading Science: Critical and Functional Perspectives on Discourse and Science* (pp. 87–113). London: Routledge.

Lim, Fei V. (2004). "Developing an integrative multisemiotic model." In K. L. O'Hallaron (Ed.), *Multimodal Discourse Analysis: Systemic Functional Perspectives.* London: Continuum.

O'Halloran, Kay L. (1999). "Interdependence, interaction and metaphor in multisemiotic texts." *Social Semiotica, 9* (3), 317–354.

O'Halloran, Kay L. (2000). "Classroom discourse in mathematics: A multi semiotic analysis." *Linguistics and Education, 10* (3), 359–388.

O'Halloran, Kay L. (2003). "Intersemiosis in mathematics and science: grammatical metaphor and semiotic metaphor." In A.-M. Simon-Vandenbergen, L. Ravelli, & M. Taverniers (Eds.), *Grammatical Metaphor: Views from Systemic Functional Linguistics* (pp. 337–365). Amsterdam and Philadelphia: Benjamins.

O'Halloran, Kay L. (2004). *Multimodal Discourse Analysis: Systemic Functional Perspectives.* London: Continuum.

O'Toole, Michael (1994). *The Language of Displayed Art*. London: Leicester University Press.

Peirce, Charles S. (1958). *The Collected Papers of C. S. Peirce*, Vol. 2. C. Hartshorne & P. Weiss (Eds.). Cambridge, MA: Harvard University Press.

Royce, Terry (1998). "Synergy on the page: exploring intersemiotic complementarity in page-based multimodal text." *JASFL Occasional Papers, 1* (1).

Saussure de, Ferdinand (1967). *Cours de linguistique générale*. R. Engler (Ed.). Wiesbaden: Otto Harrassowitz.

Sardar, Ziauddin & van Loon, Borin (2002). *Introducing Media Studies.* Cambridge: Icon Books.

Senner, Wayne M. (1989). "Theories and myths on the origins of writing: a historical overview." In W. M. Senner (Ed.), *The Origins of Writings* (pp. 1–26). Lincoln, NE: University of Nebraska Press.

Thibault, Paul (2000). "The multimodal transcription of a television advertisement: theory and practice." In A. Baldry (Ed.), *Multimodality and Multimediality in The Distance Learning Age* (pp. 311–384). Campobasso: Palladino Editore.

van Leeuwen, Theo (2000). "Some notes on visual semiosis." *Semiotica, 129* (1/4), 179–195.

van Leeuwen, Theo (2002). "Multimodality and typography." Paper presented at the *1st International Symposium on Multimodal Discourse*. Salzburg, Austria, 2002.

Chapter 4

Multimodality and empiricism

Preparing for a corpus-based approach to the study of multimodal meaning-making*

John Bateman, Judy Delin and Renate Henschel
Universities of Bremen, Germany / Leeds / Stirling, UK

Following the 'visual turn' in many areas of communication, investigators are increasingly considering explicitly both the presentation of information in forms such as photographs, diagrams, graphics, icons and so on, and interrogating their relationships with linguistically presented information. The majority of analyses currently proposed, however, remain impressionistic and difficult to verify. In this chapter we argue that the study of multimodal meaning-making needs to be placed on a more solid empirical basis in order to move on to detailed theory construction. We describe the state of the art in corpus preparation and show how this can be expanded to be of value for supporting investigative work in the area of multimodality.

1. Introduction

Following the so-called 'visual turn' in many areas of communication, it has become increasingly usual for investigators both to consider explicitly the presentation of information in forms such as photographs, diagrams, graphics, icons and so on and to place such information in combination with linguistically presented information. One of the corollaries of the broadening in the area of concern is that we are forced to deal with systems which are manifestly meaning-making (e.g. photographs, diagrams) but for which we lack the rich battery of investigative tools that we now have for linguistic entities. Whereas the application of a linguistic mode of analytic discourse is already showing significant benefits (cf. Kress & van Leeuwen 2001), the strong coupling between data and theory-construction that forms a tenet of much of modern linguistics is not yet a strong feature of 'multimodal linguistics'.

In this chapter we address this concern. We give an example where informal, interpretative claims have been made about aspects of multimodal discourse and argue that the claims demand a much more rigorous empirical basis to be taken

further. We then briefly introduce our own attempt to place multimodal study on a firmer empirical basis.

2. An example of interpretative analysis

Kress and van Leeuwen (1996) suggest that illustrated documents of a variety of kinds can meaningfully by analysed in terms of several 'signifying systems' that structure the information on the page. Of particular relevance here, is their discussion of *information value* in which they propose that the placement of elements in particular 'zones' in the visual space endows them with particular meanings. Each zone "accords specific values to the elements placed within it" (Kress & van Leeuwen 1998: 188).

While suggestive, the notion of information value as used by Kress and van Leeuwen is still in need of further clarification. Kress and van Leeuwen use it to describe oppositions between elements placed on the left of a page or image, and those placed on the right. Those on the left are considered to be 'Given'; those on the right 'New':

Given	Presented as material the reader already knows; "common sense and self-evident. . .presented as established" (Kress & van Leeuwen 1998: 189);
New	Presented as material as yet unknown to the reader; "the crucial point of the message. . .problematic, contestable, the information at issue" (Kress & van Leeuwen 1998: 189).

The analysis is appealing in that it provides a ready vocabulary for reading more out of page design than would otherwise be possible. Just as the analysis of English clauses into a Theme/Rheme structure, in which the element(s) placed at the beginning of the clause have been shown to participate with high regularity in larger text-structuring patterns (cf. Fries 1995), the Given/New patterning appears to offer a similar analytic win for the page.

But to what extent is the claim supported? Indeed, how would it be supported? The use of Given/New here is very much more abstract than that generally found in clause (or intonational unit) analyses; for Kress and van Leeuwen the Given/New in the page revolves around problematised breaks in the social norms expected. The analytic procedures for establishing to what extent this could be a reliable property of layout rather than an occasionally plausible account are unclear. Nevertheless, following on the initial presentation of the analytic scheme in van Leeuwen and Kress (1995), it has been presented again in Kress and van Leeuwen (1996, 1998) and is now itself being adopted as unproblematic, or 'Given', in some systemically-based research on multimodality (see, for example, Royce 1998; Martin 2002). Unfortunately, we have not so far found it to be supported by

designers and layout professionals in practice. It is certainly not used as a design criterion in layout. What, then, is its status?

We can see this problem particularly well in the area of newspaper design, the area within which Kress and van Leeuwen's proposal was first couched. In one of their analyses, in which they deal with a *Daily Mirror* front page (Kress & van Leeuwen 1998: 190ff.), they attribute the 'opposition' between an article about a murder on the left hand side (their 'Given' position: because we 'expect' news stories about murders and other violent activities) with a story about Michelle Pfeiffer adopting a baby on the right hand side ('New' position: famous film star acts like a mother) to Given-New organisation:

> Given, then, is the bad news: an instance of discord between lovers, with dramatic results. This is what we are exposed to every day in press reports about everyday 'private' relationships: infidelity, breakups, abuse. New is the good news. . .
>
> (Kress & van Leeuwen 1998: 190)

This is a good example of an 'impressionistic interpretative' analysis. The story told is an appealing social interpretation of a multimodal product – but it has not yet been established whether such an analysis is actually any more than a post hoc rationalisation of design decisions that occur on a page for quite other reasons.

For example, we find that the practical workflow of newspaper production would most often mitigate against a reliable allocation of the areas of the page so as to conform with the semiotic values that Kress and van Leeuwen have proposed. First of all, the relevant unit of analysis from the *production* point of view is not the page in its entirety: it is what has been termed the 'newshole' (Lie 1991) – i.e. the area that is available once advertising, mastheads, and other fixed elements have been allocated.[1] Much of the positioning on a newspaper front page is determined in advance, for example, by an a priori decision concerning where the advertising is to be and by considerations such as the need to place suitable material in the top-half of the page so that when the newspaper is folded in half and placed for sale at the newsagents enough of the newshole remains visible to sell the newspaper.

The fact that news editing and the concrete practice of newspaper production do not involve explicit conceptualisations in terms of Given/New does not, of course, mean that these categories are not employed by readers. The historical process of development in layout design may well have brought about a situation in which the semiotic values proposed by Kress and van Leeuwen hold regardless of the intentions of layout designers. But in this case, we must, on the one hand, be able to investigate readers' responses to layouts in order to provide support (or otherwise) for this interpretation and, on the other, be able to trace the development of the semiotic practise over time to see how it arose. Both investigations are scarcely possible without a tighter hold on the data that is being questioned.

We need then to ask the questions concerning the semiotic values and their realisation in layout that have been proposed by Kress and van Leeuwen more precisely. Is the entire scheme to be dismissed as a suggestive idea that did not work? Or, does the scheme apply to certain kinds of documents and not to others? Or to certain kinds of page layouts and not to others? All of these issues need to be addressed and answered as multimodal document analysis moves away from the suggestive and towards the analytic. Methods need to be adopted and documented whereby suggestive frames of analysis can be expressed as predictive and falsifiable claims about document design and meaning-making. To do this, we need to subject analyses to more detailed and systematic investigation, varying types of documents, types of consumers, types of presentation medium, and purposes so that we can get a finer grip on the meaning-making possibilities of the various semiotics in play. And to do this, we need to turn to multimodal corpora specially designed for supporting the investigation of multimodal meaning.

3. Multimodal annotated corpora

In order to provide a solid empirical basis for investigating questions of meaning-making in multimodal documents, we need to construct extensive collections of data organised in a manner that supports this inquiry. Here we draw on the experiences gained with traditional linguistic corpora. It is now part of everyday linguistic work to collect corpus instances of phenomena or patterns of concern in order to guarantee a broader and more objective basis for hypothesis formation, theory construction and verification. Moreover, the model we need for useful multimodal corpora draws particularly on *annotated* corpora – that is, collections of texts that are augmented structurally so as to support investigation of linguistic questions more readily than do simple text collections.

3.1 The origin and representation of annotated corpora

Linguistic corpora containing collections of several million words are fast becoming the norm (the British National Corpus, for example, contains 100 million words). With this mass of available 'data', it is increasingly important that the data be organised so as to support, rather than hinder, scientific inquiry.

One simple illustration of the problem here involves the phenomenon of variant linguistic forms that do not play a role in an inquiry being pursued but which make the posing of questions to a corpus more complex. If, for example, we are seeking all occurrences of the verb 'buy' in order to see what complementation patterns it occurs in, or which collocations it supports, we cannot just ask a text

collection to print out all occurrences of the string of characters 'b-u-y'. We cannot even ask it to print out all occurrences of the word "buy" – because in both cases we then do not get forms such as 'bought' and in the second case we miss forms such as "buys", "buying", etc. While relatively straightforward to avoid, such minor problems reoccur with every inquiry that one wishes to make of a corpus and easily lead to error or incomplete results.

A further illustration, a little more complex, is how to deal with a linguistic inquiry concerning uses of the modal 'can'. We can ask to retrieve all instances of the word 'can' from a corpus – but then how do we avoid all the (for this particular question irrelevant) instances of the *noun* 'can'. Again, we can do this by hand, ruling out the irrelevant cases, but this work reduces the effectiveness of using a corpus and represents a considerable overhead. More sophisticated still, if we wish to investigate the contexts in which some grammatical construction is used rather than another, then we need to be able to search for such constructions rather than particular words or sequences of words and this can be quite a difficult undertaking.

In all these cases, modern corpora provide direct support for investigation by *annotating* their contained data to include additional information that may be employed when formulating questions. That is, not only will a corpus contain the bare textual information, it will also contain information about the root form of the words used (thus enabling a single question about all occurrences of the word 'buy' in *any* of its forms), their word classes (thus enabling a question exclusively about modal 'can'), and possibly some grammatical structures or other information in addition. The provision of corpora viewed as collections of texts has largely given way to *annotated corpora*, which contain additional information for the asking of more exact linguistic questions; standard introductions to corpus linguistics describing this development in detail include Biber et al. (1998) and McEnery and Wilson (2001).

In modern annotated corpora, it is usual to employ some kind of explicit *markup language* in order to capture the extra information they contain. That is, the basic textual information is 'marked up' with the additional information to be represented drawing on standardised formats. This separates very clearly *data* from information *about that data* – which makes the information as a whole considerably easier to process and manipulate. The currently most accepted and well developed standardised formats are based on the 'Standard Generalized Markup Language': (SGML, see Bryan 1988) developed in the publishing industry and, most recently, its particular instantiation for wide-scale electronic information representation XML (the 'eXtensible Markup Language', XML). Standards for corpus annotation adopting these frameworks are also now available (cf. XCES, see CES 2000).

Both SGML and XML recommend the definition of *Document Type Descriptions* (DTDs), which specify precisely the structures that are possible in documents and the kinds of entities that can fill slots in those structures. One of the reasons for managing things in this way is that it allows documents to be automatically *checked for conformity with their intended structure*. This process is called document **validation**. It is by no means straightforward to guarantee that any sizeable collection of information is structurally correct and consistent; this is the kind of service that a DTD provides. Widely available DTD-parsers check documents for conformity with their specified DTD, so that at least formal errors may be avoided.

3.2 Annotation problems with complex data

The basic organisation of a document written in XML is very simple. Information is structured by means of **tags** in the same way as information for web pages in the hypertext markup language HTML. A piece of information is marked with a certain tag by enclosing it within an opening tag and a closing tag. If, for example, we are marking a body of text according to the XCES corpus standard as a single paragraph, we use the 'p' tag. The opening 'p' tag is written as and the closing 'p' tag as . To support a richer variety of information in the annotation, tags may also specify **attributes**. Thus, we might, for example, not only want to specify that some particular element in a corpus is a word – perhaps using the XCES tag 'w' – we may also want to give it a unique identification number, specify its part of speech information, and its root form. Providing this information makes the kinds of query problems mentioned above simple to handle. We can express this information with a complex XML mark-up such as the following:

```
<w id="J04:0230e" pos="WGv"
lemma="become">becoming</w>
```

The precise attributes that are allowed, and the kinds of values that they may take, is specified formally in the Document Type Description, and this allows this information to be formally validated for misspellings, missing brackets, wrong values of attributes, etc. Validation is already a significant reason for providing information in this structured form; we will see below, however, that many more advantages accrue from the adoption of XML.

When attempting more sophisticated linguistic annotation, the most significant problem is that of *intersecting hierarchies*. A good example of this from the area of annotation for literary editions is given by Durusau and O'Donnell (submitted).[2] One simple XCES-conformant markup of the linguistic content might break a document down into a number of identified sentences; this would use a sequence of <S> and matching closing </S> tags. Another simple XML-conformant

markup might want to indicate the division into pages that an edition employed – here we would use a sequence of <page> . . . </page> tags. Now consider an annotation for a machine-readable version of the literary work that wants to capture the page breaks *and* the linguistic divisions simultaneously. This is not straightforward simply because the linguistic division into sentences and the division into pages have no necessary relationship to one another: there is no reason why the structures imposed by the two kinds of division should embed one within the other. Thus the simplest way of capturing this information, which might appear to be something like the following:

> <page> . . . <S> This is a sentence </page> <page> that goes over two pages. </S> <S> Then there are more sentences on the page . . . </page>,

which is *not* 'legal' XML: the structures defined by the <S>-tags and the <page>-tags do not 'properly nest'. The first sentence tag is not 'closed' before its enclosing page tag is closed. Allowing such non-nesting structures would vastly complicate the machinery necessary for checking a document's conformance with its DTD .

A solution for this problem that has now established itself is that of *standoff annotation* (Thompson & McKelvie 1997). Standoff annotation recognises the independence of differing layers of annotation and separates these both from the original data and from each other. Thus, instead of having a single marked-up document where the annotations are buried within the data, the annotation information is separated off into independent annotation layers – hence the phrase 'stand off'. Each individual layer is a well-formed XML document. Contact is made with the original data *indirectly* by referring to particular elements. This solves the problem of intersecting hierarchies because within any single XML document there is no intersecting hierarchy – there is only the single hierarchy of the particular annotation layer that the document represents.

The additional technical complexity involved is that we need to be able to access the individual elements of the data in order to bind them into a variety of annotation structures. This can be achieved most simply within XML by giving each element a unique identifying label and employing *cross-references*. This is shown in a simplified example in Figure 1, where we have two annotation layers that show how a single text document is divided according to sentences and according to pages. This accepts the fact that the linguistic division into sentences and the print division into pages have no natural relationship with one another, making it inappropriate to insist that such mark-up nest properly into well-formed recursive structures simply to fulfil the SGML/XML formal restrictions. The situation illustrated takes a text where there is a page break immediately following the text: ". . . Have you, miss? Well,".

XML 'base' document

```
<w id="u-01">Have<w>
<w id="u-02">you</w>
<punc type="comma" id="u-03">,</punc>
<w id="u-04">miss</w>
<punc type="question" id="u-05">?</punc>
<w id="u-06">Well</w>
<punc type="comma" id="u-07">,</punc>
<w id="u-0B">for</w>
<w id="u-09">sure</w>
<punc type="exclamation"
     id="u-10">:</punc>
```

XML document for page breaks

```
<page id="page-01" from= "..." to="u-07"/>
<page id="page-02" from= "u-08" to= "..."/>
```

XML document for sentences

```
<s id="s-01" from="u-01" to="u-05"/>
<s id="s-02" from="u-06" to="u-10"/>
```

Figure 1. Example of standoff annotation

This information is captured by breaking the original document into a set of 'basic level' annotation units – shown in Figure 1 at the top of the figure and consisting of words ('w' tags) and punctuation ('punc' tags), each of which receives a unique identifying label as the value of their 'id' attributes. The two layers of standoff annotation shown below in the figure then refer to these labels. Thus, the first page, given its own identifying label of 'page-01', is shown as running from some base unit that we have not shown in our figure up until the unit labelled 'u-07'. The second page then runs from unit 'u-08' onwards. In a complete annotation all of the units would have received identifying labels, and so the cross-references would be complete. The other standoff layer shows precisely the same kind of information but for sentences. Each individual layer is a well-formed XML document and, because of the cross-references, there is now no problem when the distinct hierarchies fail to respect one another. This mechanism provides the basis for an open-ended set of annotation layers, each of which adds in further information to the base material.

The utility of this method relies crucially on the effectiveness of the computational software for dealing with richly structured information of this kind. The fact that the entire framework is XML-conformant is very important. The tools for writing inquiries that interrogate data structured in this way are now being refined and extended extremely quickly. This is because the main users of XML

structured data are not linguists, but standard commercial providers of information that would previously have been maintained in databases, such as sales catalogues of online companies, stock-lists, personnel data, and so on. Because of this very practical and economic demand, methods for using such data are already finding their way into the standardly available web-browsers – this virtually guarantees that it will soon be possible for XML-annotated corpora to be navigated and manipulated using widely available and familiar tools rather than complex, corpus-specific schemes and software.

4. The GeM model: Layering for classification

In this section, we set out how we are approaching the design of multimodal corpora drawing on the state of the art for annotated linguistic corpora described in the previous section. We have been pursuing these aims in the context of a research project, the 'Genre and Multimodality' project GeM (http://www.purl.org/net/gem).[3] The basic aim of GeM is to investigate the appropriateness of a multimodal view of 'genre': that is, we are seeking to establish empirically the extent to which there is a systematic and regular relationship between different **document genres** and their potential realisational forms in combinations of text, layout, graphics, pictures and diagrams. More detailed introductions to the GeM model and its motivation can be found in Delin et al. (2002/3) and Delin and Bateman (2002).

4.1 The GeM model

Our starting point for considering genre draws primarily on linguistic uses, such as evident, for example, in Biber (1989) or Swales (1990). We also emphasise and build on the social 'embeddedness' of genres: texts look different because they are to function in different social contexts (cf. Halliday 1978; Martin 1992). Moreover, as a final step, we then reconnect this notion to the practical contexts of production and consumption of the discussed genres; that is, genres are also partially defined by their 'rituals of use' and the application of various technologies in the construction of their members (cf. Kress & van Leeuwen 2001).

The first attempt that we are aware of that provided a detailed model of multimodal genre was that of Waller (1987). It took into consideration the vital contributions of language, document content, and visual appearance as well as practical conditions of production and consumption. Our own work draws upon and extends this framework by examining the interdependencies between possible characterisations of genre on the one hand and of the various functional

constraints on the other. The basic levels of analysis that the project has defined are then as follows:

1. *Content structure:* the 'raw' data out of which documents are constructed;
2. *Rhetorical structure*: the rhetorical relationships between content elements; how the content is 'argued';
3. *Layout structure:* the nature, appearance and position of communicative elements on the page;
4. *Navigation structure*: the ways in which the intended mode(s) of consumption of the document is/are supported; and
5. *Linguistic structure:* the structure of the language used to realise the layout elements.

We suggest that document genre is constituted both in terms of levels of description such as these, and in terms of constraints that operate during the creation of a document. Document design, then, arises out of the necessity to satisfy communicative goals at the five levels presented above, while simultaneously addressing a number of potentially competing and/or overlapping constraints drawn from:

Canvas constraints:Constraints arising out of the physical nature of the object being produced; e.g.: paper or screen size; fold geometry such as for a leaflet; number of pages available for a particular topic.

Production constraints: Constraints arising out of the production technology: e.g. limit on page numbers, colours, size of included graphics, availability of photographs; and constraints arising from the micro-and macro-economy of time or materials: e.g. deadlines; expense of using colour; necessity of incorporating advertising.

Consumption constraints: Constraints arising out of the time, place, and manner of acquiring and consuming the document, such as method of selection at purchase point, or web browser sophistication and the changes it will make on downloading; also constraints arising out of the degree to which the document must be easy to read, understand, or otherwise use; fitness in relation to task (Read straight through? Quick reference?); assumptions of expertise of reader, etc.

Particular genres are constituted by regularly recurrent and stable selections and particular sets of constraint satisfactions. And these can only be ascertained *empirically* by the investigation of a range of document types.

4.2 Designing and populating a multimodal corpus

We have already seen the basic technological requirements sufficient for constructing a multimodal corpus. When we adopt the GeM layers of analysis, it is possible to consider each one as a single layer of standoff annotation just as was illustrated

for the simple page and sentence example of Figure 1. This has now been done with Document Type Descriptions specified in XML-form for each layer. As usual with formalisation, the demand for complete specification has resulted in a considerable number of refinements to the original model we have just sketched. These are set out in full in the technical documentation for the corpus design (cf. Henschel 2002). Here we focus on just one layer of annotation, the layout structure, which has been developed within the GeM project. For the purposes of this chapter we will also concentrate on the addition of *pages* involving multimodal content rather than go into the details of considering entire documents.

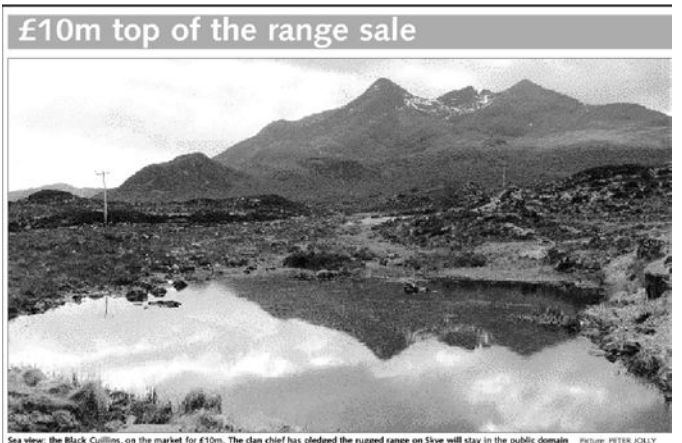
As we have seen, a precondition for standoff annotation is to establish a single document containing the marked-up 'basic units' of any document being added to the corpus. With GeM, these base level units range over textual, graphical and layout elements and give a comprehensive account of the material on the page, i.e. they comprise everything which can be seen on the page/pages of the document. The base units we define for GeM include: orthographic sentences, sentence fragments initiating a list, headings, photos, drawings, figures (without caption), captions of photos, text in pictures, icons, table cells, list headers, page numbers, footnotes (without footnote label), footnote labels, and so on. Each such element is marked as a base unit and receives a unique base unit identifier. The base units provide the basic vocabulary of the page – the units out of which all meaning-carrying configurations on the page must be constructed.

Details concerning the form and content of each base unit are not represented at this level. All such information is expressed in terms of pointers to the relevant units of the base level from the other layers of annotation. As suggested above, this standoff approach to annotation readily supports the necessary range of intersecting, overlapping hierarchical structures commonly found in even the simplest documents. Single base units are commonly cross-classified to capture their multifunctionality and can, for example, contribute to a visually realised layout element as well as simultaneously functioning as a component of a rhetorical argument. This ensures that we can maintain the logical independence of the layers considered.

Thus, to take a relatively simple example, if we were annotating the part of a page shown in the lower part in Figure 2, we would construct a base document along the lines of the XML annotation shown at the top of the figure.[4]

Each typographically distinct element on the page is allocated to a different base unit. The first unit (identified by the label 'u-01') corresponds to the headline at the top of the page extract; here we can see that the only information captured here is the raw text "£10m top of the range sale" – typographical information, placement on the page, rhetorical function (if any), etc., are not represented. The second unit does the same job for the large photograph – the 'raw picture' is repre-

```
<unit id="u-01">£10m top of the range sale</unit>
<unit id="u-02" src="cuillins-pic.jpg" />
<unit id="u-03">Sea view:</unit>
<unit id="u-04">The Black Cuillins, on the market for £10m. </unit>
<unit id="u-05">The clan has pledged the rugged range on Skye will
stay in the public domain</unit>
<unit id="u-06">Picture:</unit>
<unit id="u-07">Peter Jolly</unit>
<unit id="u-08">Clan chief puts Black Cuillins on the market </unit>
<unit id="u-09">Raymond Duncan</unit>
<unit id="u-10">One of Scotland's ...</unit>
...
<unit id="u-70"> Inside</unit>
<unit id="u-71">The Chief's Castle</unit>
<unit id="u-90">Page 2</unit>
<unit id="u-91" alt="line" />
...
<unit id="u-99">Continued on page 2</unit>
```



£10m top of the range sale

Sea view: the Black Cuillins, on the market for £10m. The clan chief has pledged the rugged range on Skye will stay in the public domain Picture: PETER JOLLY

Clan chief puts Black Cuillins on the market

RAYMOND DUNCAN

ONE of Scotland's most famous mountain ranges has been offered for sale at more than £10m by a clan chief.

The aim is to raise funds to repair an 800-year-old castle which is the seat of a leading clan.

Skye's Black Cuillins, one of the country's most spectacular landscapes and an area of major international importance, has been put on the market by the 29th Chief of the Clan MacLeod.

John MacLeod of MacLeod yesterday described the decision to sell as the most difficult of his life and "an extremely painful experience".

The move brought strong criticism from Western Isles MP Calum MacDonald. He said: "The idea of selling the Cuillins is ridiculous but at such a price it's obscene. MacLeod should hang his head in shame for trying to exploit what God has given the people of Skye. It goes to show that one thing worse than a thieving foreign landlord is a greedy Scottish one."

He also said it demonstrated the need to get land reform through Parliament at top speed to stop "the obscenity of landlords trying to cash in on public concern for areas like the Cuillins".

The 23,300 acres include more than 14 miles of coastline, two salmon rivers, a sheep farm and farmhouse.

News of the sale prompted a call from Highland Council for the mountains to be bought for the nation. It wants the "world treasure" bought by a partnership of public bodies, tenant farmers and local communities accessing Heritage Lottery Funding at District Valuer's valuation.

Potential partners could include the council, Highlands and Islands Enterprise, Scottish Natural Heritage and the National Trust for Scotland.

Councillor Michael Foxley, chair of the land and environment select committee, said: "We will be seeking a meeting in the very near future to set the ball rolling."

The 64-year-old owner of Dunvegan Castle said the move to sell the 35 square miles of rugged landscape was principally economic. Mr MacLeod called the mountain range "part of my soul". He vowed the Cuillins, once a temporary refuge for Bonnie Prince Charlie and now a paradise for geologists, botanists and mountaineers.

Continued on Page 2

INSIDE

The chief's castle – Page 2

Figure 2. Page extract from a newspaper and corresponding base unit annotation

sented indirectly by a link to a source file containing the image ('cuillins-pic.jpg') just as is done in HTML files for web presentation. The next five units describe the caption(s) underneath the picture; 'u-03' is an introductory label for the caption "Sea view:", 'u-04' and 'u-05' are two 'sentence'-like units making up the body of the caption, and 'u-06' and 'u-07' give information about the photographer. Again, the only role played by this division into units is to provide labels that subsequent layers of annotation can call on by cross-references when describing their functions on the page. Even the fact that the units are approximately ordered following their vertical ordering on the page is not significant – they could in fact be written in any order. This means that any collection of such units can be picked out by the other layers of annotation in order to carry differing functions as necessary.

In contrast to the simplicity of the base layer, the other annotation layers are rather more complex. The layout layer of annotation is probably the most complex, however, in that it has several different tasks to perform in capturing the layout decisions taken in a page. These may be summarised as follows. The layout structure must:

a. capture all the particular typographical distinctions drawn on the page – such as, for example, the fact that certain elements are entirely in capitals, others are in bold, some are in one type face and others in another, and so on;
b. represent the visually expressed hierarchy of related 'blocks' on the page – such as, for example, the relative grouping of a picture with its caption as a unit with respect to some other visual element for which the picture-plus-caption functions as elaborating material;
c. relate the visual hierarchy of layout blocks to their concrete positions on a page of information.

Each of these kinds of information is managed as a locally complete XML structure. We show all three briefly for the selected newspaper fragment.

The 'backbone' of the layout annotation is provided by the second of these kinds of information: the visually oriented hierarchy of layout elements. This is determined by a set of methodological heuristics for decomposing the information on the page. One such heuristic is for the analyst to consider the relative visual prominence or salience of the blocks on the page. This can be supported by a range of 'tricks': for example, by progressively reducing the resolution of the image when displayed. The blocks which dissolve first are the lowest in the layout unit hierarchy (e.g. the smallest typographically displayed letters and words), those that dissolve into each other last are the highest level units of the hierarchy. A second heuristic is to consider which chunks of information 'belong together' – i.e. if one block were to be 'moved' on the page, which others are 'drawn along' with it. For example, if we were to move the photograph on our page, then it is natural that the caption

would be drawn with it, and less likely that the body of the text or the headline immediately move: although there would be limits to this in the context of the page as a whole as the individual units making up this 'story' would not like to be separated. General proximity is thus to be maintained, which is itself an argument for maintaining all the units shown as a single higher-level layout unit.

Furthermore, within this, the block in the middle of the second column of text stating that more information (of a particular kind: i.e. 'the Chief's castle') exists and providing navigation information about where that information is located ('inside' and 'Page 2') can also be moved relatively freely within its enclosing text block, arguing for its treatment as a distinct layout unit at an intermediate level in the overall hierarchy. An example of this kind of structuring is shown in Figure 3.

In general, the hierarchical structures proposed should be conservative – that is, when there is no strong evidence in favour of a strict hierarchical relationship, we prefer to posit a flat structure rather than insisting on some particular hierarchicalisation. The layout hierarchy captures dependency relationships between visually discovered elements on the page but no longer includes information about the precise physical location of those elements on the page. It is therefore a significant abstraction away from the source document and generalises over a set of 'congruent' possible realisations.

A layout hierarchy is represented as a simple nested XML structure made up of 'layout chunks' and 'layout leaves'. Layout chunks can have further layout chunks embedded within them to set up the recursivity of the structures represented. Terminal elements in the structure are represented as layout leaves. Each such unit again receives its own unique identifying label and the entire structure is placed within a single enclosing XML tag called the 'layout root'. The contents of each layout unit, that is, the elements on the page that comprise them, are identified in the way standard for standoff annotation, i.e. the layout leaves contain cross-references to the identifiers of the corresponding base units. The layout structure corresponding to the example in Figure 3 is then as shown in Figure 4.

The interested reader can follow through the structure and the cross-references as identified in the base units of Figure 2 to confirm that the hierarchical view thus created does indeed correspond to the hierarchy given in Figure 3. This should help make it clear why proper computational tools for checking the formal consistency (e.g. are all the identifying labels used actually defined somewhere?) are so important.

The representation of the orthographic and typographic information is then relatively simple. A set of XML-specifications state which *layout units* have which typographical features. In this way, it is straightforward to make generalisations over subhierarchies drawn from the layout structure. For example, all the layout units corresponding to a block of text that is realised uniformly in terms of its

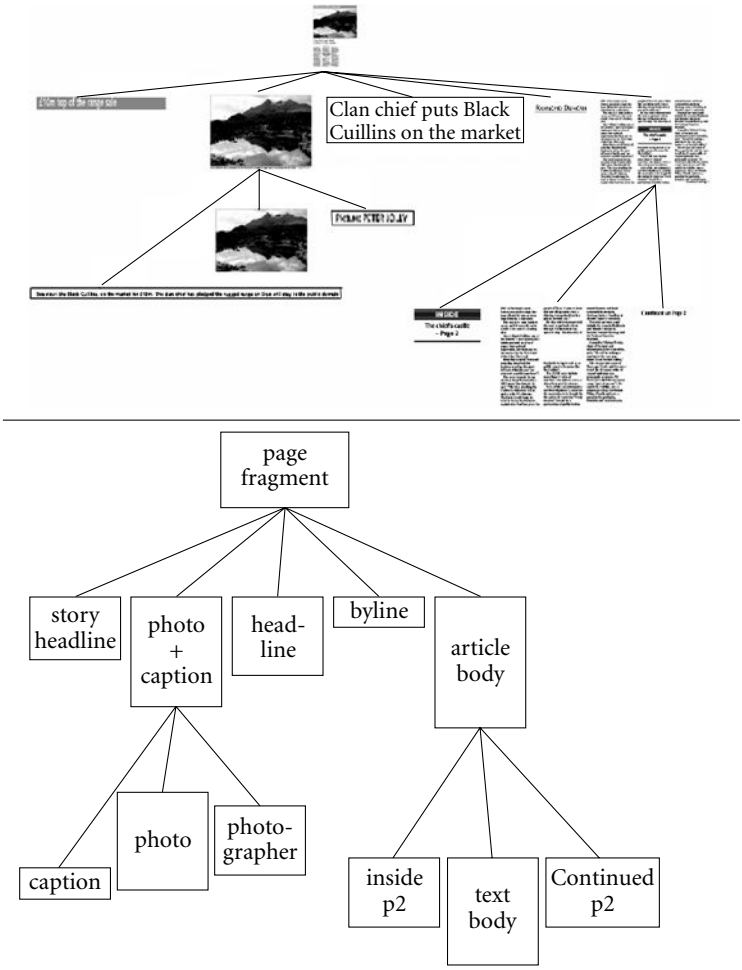


Figure 3. Derivation of hierarchical layout structure from the example page

typography may be grouped as a single node in the layout structure, and it is this node which has the corresponding typographic features associated with it. This allows information to be expressed concisely without repetition.

There are already very extensive vocabularies for describing typographical features: we adopt these for this aspect of the GeM annotation scheme rather than develop a further, ad hoc set of terms. Concretely, we use the typographical distinctions described as part of the XML formatting objects standard. An example of such a specification for the unit corresponding to the headline at the top of the page is given in Figure 5.

```
<layout-root id="lay-01">
   <layout-leaf id="lay-02" xref="u-01"/>
   <layout-chunk id="lay-03">
      <layout-leaf id="lay-04" xref="u-03 u-04 u-05"/>
      <layout-leaf id="lay-05" xref="u-02" />
      <layout-leaf id="lay-06" xref="u-06 u-07" />
   </layout-chunk>
   <layout-leaf id="lay-07" xref="u-08" />
   <layout-leaf id="lay-08" xref="u-09" />
   <layout-chunk id="lay-09">
      <layout-leaf id="lay-10" xref="u-70 u-71 u-90"/>
      <layout-leaf id="lay-11" xref="u-10 ... " />
      <layout-leaf id="lay-12" xref="u-99" />
   </layout-chunk>
</layout-root>
```

Figure 4. Layout structure represented in XML according to the GeM scheme

```
<text xref="lay-02"              font-family="sans-serif"
   font-size="18"                font-style="normal"
   font-weight="bold"            case="mixed"
   justification="left"          color="white"
   background-color="grey"/>
```

Figure 5. Typographical headline annotation

The final component of the layout annotation layer adds in the information about precise placement within a page. We separate a general statement of the *potential* placement strategy employed on a page from that of the hierarchical layout structure for that page. Placement is then indicated by adding to the layout elements an 'address' given in terms of the general positions defined possible for their page. We have found this separation of information to be worthwhile for a number of reasons. First, it is quite possible that minor variations in the precise placement of layout elements can be undertaken for genre-specific reasons without altering the hierarchical relationships present. Second, the separation of placement information makes it possible to state generalisations over the physical placement that are inconveniently expressed at the level of individual layout elements: for example, it is common that pages use various alignments for their material – this alignment can hold over portions of the layout structure that are not strongly related hierarchically. Good illustrations of the consequences of varying such alignments or non-alignments are given, for example, in Schriver (1997: 314) for complex instructional texts.

In order to fully capture these possible dimensions of variation, we express within-page placement in terms of an **area model**. Area models divide the space on a page into a set of hierarchically nested **grids** or tables. Since the grid technique is

one that is commonly employed in professional design, it is often straightforward and useful to allow this information to be expressed directly in our annotation; this is particularly the case for newspapers, which are traditionally prepared and designed using pages divided into columns. However, in contrast to the generic column-structuring of newspapers, the function of the area model is more specific in that it provides particular physical reference points for the defined layout elements. Layout elements from the layout structure are then placed in correspondence with particular elements drawn from the page's grid structure. This is necessary because simply stating that some layout unit divides, for example, into three sub-elements still leaves very many options open for those sub-elements' physical placement, both within the general space defined by their parent layout unit and with respect to one another.

The grid structure of the area model for our example page extract is shown in Figure 6. Here we can see that the main body of the page is annotated as having a 'row' structure rather than a full grid. Some of these rows are themselves subdivided into further row or column structures.

This kind of area model is quite characteristic for newspapers both with respect to the use of an overall column structure, which is picked up as columns of various sub-areas, and with respect to the relatively frequent use of 'insets', which relatively arbitrarily 'cover' parts of the grid structure so that it is no longer avail-

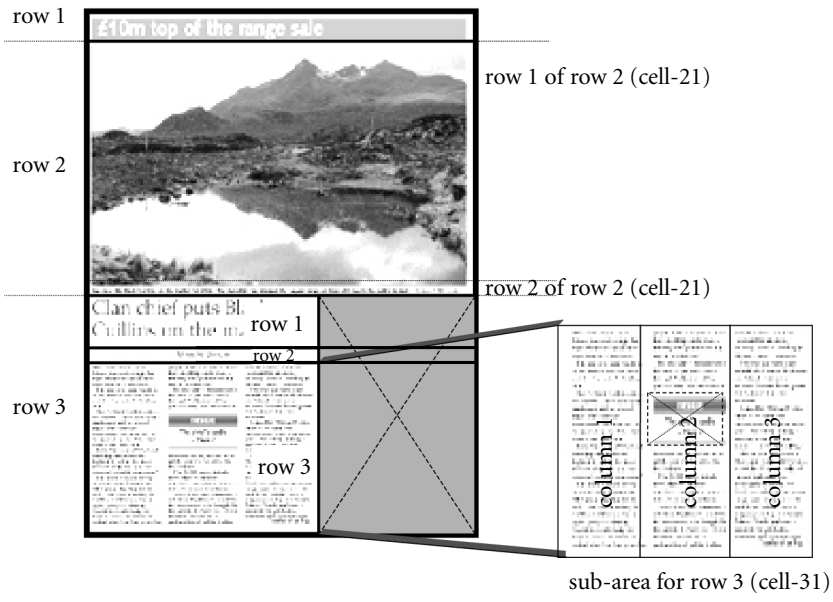


Figure 6. Area model represented as a grid structure for the page

able for some particular content. This is commonly the case for advertisements and other rhetorically distinct information such as the navigation elements in the middle of column 2 of the sub-area of row 3 within top-level row 3.[5]

Although there are many interesting further issues that arise with this layer of annotation, space precludes their discussion here. Readers are referred to the GeM technical documentation for a more complete account. All of the pages of the documents being added to the GeM corpus are described in the general terms that have been set out here.

Providing annotation layers as described in this section for all of the GeM layers is then the main task involved in constructing a multimodal corpus of this sort. We use XML so that we can rely on standard tools and techniques for storing the data, checking their integrity, and for presenting various views of the data when considering analysis. This then places multimodal corpus design for the kinds of documents that we are considering on a firm technological foundation. We also use XML, however, to be able to make use of the tools that are now emerging in the structured data representation industry for presenting queries and for searching for regularities in the data captured. And it is to this that we now turn.

5. Examples of using a GeM-annotated corpus

Space precludes anything here but a single brief example of using the GeM-annotated corpus for linguistic research – drawing on our example in Section 2. Although the corpus needs to be considerably extended in coverage before we can approach the kind of statements now possible in linguistic corpus analysis, we nevertheless believe that the approach outlined represents a sound methodological direction for eventually achieving this goal. Our discussion in this section must therefore be seen as merely suggestive of the possibilities that open up when multimodal corpora are available in the form we propose.

We have made much of the fact that we now have a method and framework for adding multimodal pages into a corpus of multimodal documents that is richly annotated and XML-conformant. A prime motivation for this direction is to be able to avail ourselves of another area of the emerging XML industry: that is the area of *searching and manipulating* XML documents. In essence, the only reason to put the effort into the highly structured forms of representation necessary for a representation such as XML is the promise of being able to get out more than one has put in. In the case of linguistic corpora, we are seeking the ability to ask questions of our corpus in sufficiently flexible and powerful ways as to promote theory construction and testing.

The components of the XML standard that are relevant here are those concerned with finding selected elements within a set of XML-structured data. One large-scale effort in the World-Wide Web community that is concerned with this task is the 'XPath' group. This group has formulated an approach to finding elements within an XML structure by specifying in a very general way 'paths' from the root of the XML structure to the element that is being sought. The path is similar to that used for files or folders on a computer system: as elements in XML may be recursively structured, and each structural element is identified by its tag, this provides a ready addressing mechanism to navigate around XML structures of arbitrary size and complexity. As a simple example, if we wanted to locate within a layout structure the top level layout chunks, then all we need write is an XPath specification such as:

```
/layout-root/layout-chunk
```

and the result, when passed to a standard XPath-processor, would be the set of layout-chunks immediately embedded within the layout-root. A variety of further constructions make the XPath specifications into a powerful way of locating sets of parts of XML documents that conform to given requirements – which is exactly what is needed for corpus investigation.

For example, the following applied to the representation for a linguistic corpus that we suggested in Section 3.2. above would return the contents of all elements tagged as w-elements without any annotation – i.e, just the words.

```
//w
```

More indicative of the power of the XPath mechanism is the following, which would give us all instances from the corpus where a word has been classified as having the part of speech designated "WGv" (by means of the value of the 'pos' attribute):

```
//w[@pos="WGv"]
```

Further constructs allow us to sort these, again according to various criteria, or to impose further restrictions (e.g. all such words that follow or precede some other class).

The XPath language is being defined and implemented independently of any linguistic concerns – it is again subject to the primarily economic demands that are also driving the development of XML. The fact that we can immediately use the results of this development for linguistic work is solely because of the XML-conformant nature of our annotation scheme.

Returning then to the question of the distribution of Given/New material on the page, as analysed by Kress and van Leeuwen in Section 2, we can now design

a series of empirical and corpus-based studies for its investigation. If their framework were to be established as correct, then a news story placed on the left of the page is *by virtue of that placement* inherently 'Given' with respect to, or relative to, a story that is placed on the right of the page. Several experimental setups can be envisaged for investigating this claim. We might ask readers to rate the various stories and their pictures on a newspaper front page on a scale running from 'expected' to 'exceptional' and then see if there is any correlation with page placement. Alternatively, we might select articles that are on the 'left' of the page and those on the 'right' (allowing for area model and canvas perturbations) and have readers judge these with respect to one another. Then we might 're-generate' newspaper front pages with the articles on the left and those on the right swapped to see if readers' judgements are effected.

For all of these tasks, we can profitably employ an appropriately annotated corpus of newspaper front pages. The selection of items on the left and those on the right probably needs to be made with some sensitivity to the generic layout of pages: it might be that we need to filter out the advertisements, or the table of contents, that regularly happen in some newspaper to occupy the leftmost (or rightmost) column. This can be pursued by following through the rhetorical structure annotation of the page, finding the main nuclear elements, following the cross-references back to the involved base-units, and selecting just those that are positioned in the layout structure to the right or to the left of the corresponding area models. This is exactly the kind of manipulation for which the XML component XPath is being designed. We might also need to separate out experimental runs involving pages with very different general layout schemes – for example, those which are predominantly vertically organised and those which show a horizontal organisation; again these kinds of properties can be calculated and made into an explicit selection criterion on the basis of the area model.

Asking readers to judge the articles for degrees of Given/New can also be seen as an annotation task: and this can be supported by existing annotation tools for XML. To run our experiment, we might then define an additional 'experimental' layer of XML markup in which experimental subjects choose a rating for presented parts of a page or of selected articles shown independently of their position on a page. The selection of the articles is itself straightforward in that once we find the set of base units that constitute an article, we simply present these as a running text, or text with pictures, ignoring the other information given in the layout structure of the page. Our experimental layer of annotation then associates these articles with Given/New ratings in senses hopefully including the very abstract ones intended by Kress and van Leeuwen. We then run over the resulting annotations, displaying the actual page placements of the articles with specific ratings. If the given/new claim of Kress and van Leeuwen is correct, then we should see

clear preferences emerging. There may, however, be additional variables to take into consideration.

We do not yet know what the outcomes to experiments such as these would be, but given a sufficiently broad GeM-annotated corpus the experiments themselves will be far simpler to run since the preparation of experimental materials is considerably facilitated. The fact that we shall probably obtain clues for further more refined hypotheses which will in turn require further experiments, with further materials to be prepared, is another strong motivation for automating as much of the materials preparation as we can. And this can only be done with a corpus which is annotated in a way similar to that proposed here.

6. Conclusions and directions for the future

We have argued that it is essential that multimodal analysis that draws on linguistic methods of analysis adopt a more explicit orientation to corpora of organised data. Only in this way is there a hope of demonstrating that certain, currently more impressionistic styles of analysis in fact hold germs of truth (or otherwise). By presenting a first view of an analytic framework for organising multimodal (page-based) data, we have tried to show how this can be done. The availability of increasingly large-scale and inclusive bodies of such data should enable work on multimodal analysis to shift its *own* genre – we expect that the kinds of discourse adopted in analyses of this kind will be able to draw nearer to empirical linguistic discourse and to go beyond styles of discourse more closely allied with literary or cultural analysis.

While it may turn out that the kinds of meaning-making involved in multimodal discourse are not amenable to analysis in this way, that the role of the interpretative subject is too great and the constraints on meaning brought by the products analysed too weak, we see it as at least methodologically desirable that we pursue this path before dismissing it.

We believe the current layers of the GeM model to be the minimum necessary for capturing the basic semiotic meaning-making potential of multimodal pages. They are also, however, clearly not sufficient for all that one needs to ask – for example, we have deliberately left out the detailed annotation of the *contents* of pictorially realised elements of pages. This is one reason why the annotation scheme has been defined in a manner which is deliberately open-ended in terms of the information it covers. Further layers of annotation need to be considered. One obvious candidate for such a layer is the detailed analytic scheme proposed by Kress and van Leeuwen (1996). In addition, although we have said very little about those levels of meaning-making which are more usually of concern to lin-

guists: i.e. the linguistic structure, we believe that the form of annotation presented here articulates well with the kind of linguistic analysis that is capable of representing the rich connections between language forms and their underlying functions, and that the model as a whole then forms the most sophisticated attempt to model explicitly all the layers that constitute genre available to date.

Clearly, after setting out the motivation and methods for this approach to multimodal corpora construction, the main body of work remains to be done. Only when we have such corpora can we start putting the programmes of exploration sketched in the previous section into action. That is a considerable and long-term task; where it will take us in our understanding of the meaning-making potential of multimodal documents is something that only the future will tell.

Notes

* The GeM project was funded by the British Economic and Social Research Council, whose support we gratefully acknowledge.

1. In other genres, the area conceived of as available for layout may not be a page at all: it may be a spread, a run of pages, or a screenful.

2. Durusau and O'Donnell's example is actually rather more complicated. They also give an excellent overview of possible approaches and problems.

3. 'Genre and Multimodality: a computer model of genre in document layout'. Funded by the British ESRC, grant no. R000238063. Project website: http://www.purl.org/net/gem

4. This page extract is selected from the front page of an edition of the Scottish daily newspaper, *The Herald*. It is reproduced by permission.

5. Note that to describe what is going on in the case of the newspaper page fully, we have an interesting interaction between several other layers of the GeM model. The fact that a newspaper page is organised throughout in terms of columns is nowadays one of the *canvas* constraints that hold for the genre: no matter how the individual articles are organised in terms of their own area models, they must be 'poured' into the mould provided by the canvas, which, for newspapers, consists of columns. In earlier times, when print technology was more restrictive, we can even imagine the 'column nature' of newspapers being a *production* constraint – i.e. one imposed by the technology of production and so not variable for different purposes. The GeM constraints form a natural hierarchy; for example, canvas constraints can only be varied within the range of possibilities that the production constraints provide for.

References

Biber, Douglas (1989). "A typology of English texts." *Linguistics, 27*, 3–43.

Biber, Douglas, Conrad, Susan, & Reppen, Randi (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Bryan, Martin (1988). *SGML: An Author's Guide to the Standard Generalized Markup Language*. New York: Addison-Wesley Publishing Company.

CES (Corpus Encoding Standard). (2000). "Corpus encoding standard. Version 1.5." http://www.cs.vassar.edu/CES

Delin, Judy, Bateman, John, & Allen, Patrick (2002/3). "A model of genre in document layout." *Information Design Journal, 11* (1), 54–66.

Delin, Judy & Bateman, John (2002). "Describing and critiquing multimodal documents." *Document Design, 3* (2), 140–155.

Durusau, Patrick & O'Donnell, Matthew B. (submitted). "Implementing concurrent markup in XML." *Markup Languages: Theory and Practice*.

Fries, Peter H. (1995). "Themes, methods of development, and texts." In R. Hasan & P. Fries (Eds.), *On Subject and Theme: A Discourse Functional Perspective* (pp. 317–360). Amsterdam: Benjamins.

Halliday, M. A. K. (1978). *Language as Social Semiotic.* London: Edward Arnold.

Henschel, Renate (2002). *GeM Annotation Manual, Bremen and Stirling*. University of Bremen and University of Stirling. http://www.purl.org/net/gem

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images: the Grammar of Visual Design.* London and New York: Routledge.

Kress, Gunther & van Leeuwen, Theo (1998). "Front pages: The (critical) analysis of newspaper layout." In A. Bell & P. Garrett (Eds.), *Approaches to Media Discourse* (pp. 186–219). Oxford: Blackwell.

Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication.* London: Arnold.

Lie, Hakon K. (1991). The Electronic Broadsheet: All the News That Fits the Display. M.A. thesis. Boston: School of Architecture and Planning, MIT. http://www.bilkent.edu.tr/pub/WWW/People/howcome/TEB/www/hwl_th_1.html

Martin, James R. (1992). *English Text: System and Structure.* Amsterdam: Benjamins.

Martin, James R. (2002). "Fair trade: negotiating meaning in multimodal texts." In P. Coppock (Ed.), *The Semiotics of Writing: Transdisciplinary Perspectives on the Technology of Writing* (pp. 311–338). Turnhout: Brepols.

McEnery, Tony & Wilson, Andrew (2001). *Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Royce, Terry D. (1998). "Synergy on the page: exploring intersemiotic complementarity in page-based multimodal text." *Japan Association for Systemic Functional Linguistics* (*JASFL*) *Occasional Papers, 1*, 25–49.

Schriver, Karen A. (1997). *Dynamics in Document Design: Creating Texts for Readers.* New York: John Wiley and Sons.

Swales, John M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Thompson, Henry S. & McKelvie, David (1997). "Hyperlink semantics for standoff markup of read-only documents." In *Proceedings of SGML Europe'97*.

van Leeuwen, Theo & Kress, Gunther (1995). "Critical layout analysis." *Internationale Schulbuchforschung, 17*, 25–43.

Waller, Robert (1987). The Typographical Contribution to Language: towards a Model of Typographic Genres and Their Underlying Structures. PhD dissertation. Department of Typography and Graphic Communication, University of Reading, Reading, UK.

Part II

Analyses and applications

Chapter 5

On the effectiveness of mathematics*

Kay L. O'Halloran
National University of Singapore

In this chapter I propose that viewing mathematics as a multisemiotic construction consisting of language, mathematical symbolism, and visual display may prove productive in two different ways. Firstly, this approach permits examination of the nature of mathematics itself, and secondly, it allows consideration of the evolution of mathematical discourse and its relation to the sciences. In the first case, an integral factor in the 'success' of mathematics appears to be the use of the three semiotic resources with specific grammars which permit semantic expansions in limited domains through the available meaning potential of each resource. Especially significant is the development of the grammar of mathematical symbolism which bridges visual and verbal descriptions. Further to this, the use of multiple semiotic resources makes it possible that semiotic metaphors involving semantic shifts as functional elements are reconstrued in another semiotic. The question of the success of mathematics in the natural sciences should also be considered in a historical context. Mathematics initially developed as a specialised multisemiotic discourse to describe our physical experience of the world (for example, arithmetic, geometry and calculus). Although from these origins mathematics has since evolved to other levels of abstraction, it nonetheless continues to be useful in the natural sciences because it was specifically developed to provide the semiotic tools for that construction of reality. In other words, mathematics provides the means for what has since evolved into physical, biological and social sciences. Seen in this light, it is understandable that mathematics is effective. Mathematics simultaneously defines, enhances and constrains our understanding of the world in scientific terms.

1. Introduction

Wigner's (1960) essay *The Unreasonable Effectiveness of Mathematics in the Natural Sciences* resulted in much debate from mathematicians, philosophers and scientists in what were essentially attempts to explain the usefulness of mathematics in the natural sciences. Although various approaches have been offered with varying degrees of success (for example, Davis & Hersh 1986; Mickens 1990) there has been

an absence of a semiotically informed response to the question of how and why mathematics is effective in describing the physical world. In this chapter I attempt such a response as previously advocated by others.[1]

Conceptualising mathematics as a multi-semiotic discourse involving language, mathematical symbolism and visual display is a useful approach as it permits examination of the meaning potentials of the three resources made available through their respective grammars. Grammar is seen as a series of system networks which have various semantic options (the abstract system), and the mathematics text (the actual written text) is the result of particular choices from those available network options. Close textual analysis reveals how the three resources function both individually and also together as a tool to create meaning in mathematics. To date such an account of mathematical discourse does not exist (Rotman 2000: 42). This approach shows how mathematics simultaneously defines, enhances and constrains our understanding of the world in scientific terms.

The question of the success of mathematics in the natural sciences is also considered from a historical perspective in order to understand some of the forces which shaped the design of the discourse. It appears that mathematics originally developed as a specialised multisemiotic discourse to construct aspects of our physical experience of the world. Although from these origins mathematics has since evolved to other forms of abstraction, the discourse perhaps nonetheless continues to be useful in the natural sciences because it was *specifically* developed to provide the semiotic tools for that construction. Mathematics provides the semiotic means for description, prediction and prescription in what has since evolved into the physical, biological and social sciences. Seen in this light, it is understandable that mathematics is effective, though there are limits to this effectiveness as there are limits to what can be understood scientifically.

I attempt to substantiate a social semiotic approach in the following discussion and in doing so demonstrate how a systemic functional perspective of mathematics can contribute to a useful understanding of mathematical and scientific discourse. The strategy adopted here is to provide a brief glimpse of the historical development of mathematics during the Renaissance as seen in various pictures and diagrams. Very often these dated images seem to appear in mathematics textbooks for interpersonal reasons (to engage the reader) rather than for experiential meaning (that is, content). The images are often presented as exotic historical relics which function as points of interest for potentially disinterested students. However, visual images do provide some way of making accessible to a more general audience the historical unfolding of the functions, semantic selections and meanings made in mathematics. The Renaissance images presented in this chapter are linked to a more general discussion of the nature of the evolution of mathematical symbolism and language during the beginnings of modern science in the

1600–1700s. As part of this discussion, an outline of the philosophy which shaped the direction of modern mathematics and science is given. From the nature of this project we can begin to appreciate what modern mathematics was originally designed to do, and the means through which this was achieved. Following this discussion, I briefly discuss the dimensions of meaning found in modern mathematics before returning to the question poised by Wigner. My own position is that it is not perhaps surprising that mathematics is effective in describing our physical world. The degree of success of that description, however, is a separate question.

2. Wigner on the effectiveness of mathematics

The success of mathematical descriptions in the sciences has apparently attracted much interest. "Clearly mathematics is a very effective tool/language for the formulation of scientific theories. Why is this the case?" asks Mickens (1990: v). The framing of this perception of mathematics and its seemingly perplexing relation to science was shaped (at least in part) by Wigner (1960):

> The first point is that the enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious and that there is no rational explanation for it. Second it is just this uncanny usefulness of mathematical concepts that raises the question of the uniqueness of our physical theories.
>
> (Wigner 1960: 293)

> The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure even though perhaps also to our bafflement, to wide branches of learning.
>
> (Wigner 1960: 306)

Mickens (1990) is a collection of eighteen responses to Wigner some thirty years after the essay first appeared. The responses are mixed formulations from various theoretical viewpoints. However, despite the various approaches adopted in these papers, Mickens claims that the problem of the effectiveness (reasonable or not) of mathematics in the sciences is not settled. Oldershaw (1990: 137) summarises the situation: "With notable exceptions, many of the previous expositions on the math-science relationship have been hampered by intrusions of obscurantism, anecdotal reasoning, and avoidance of definite conclusions."

While the mathematics/science relationship may never be resolved, respondents such as Oldershaw (1990: 145) question Wigner's rhetorical style and his mystification of the issue: "Perhaps we sometimes romanticise and mysticise mathematics and mathematical physics in order to reassure ourselves that we are doing

something quite extraordinary." Oldershaw's view partially informs the position adopted here where the limits of the contribution of mathematics to our understanding of the physical world is opened up for further thought.

Despite concluding that his reasons are not quite sufficient to explain the usefulness of mathematics, Hamming (1980) would appear to support the historical semiotic approach adopted here as he claims that (1) We see what we look for; (2) We select the kind of mathematics to use; and (3) Science in fact answers comparatively few problems. Hamming's (1980) three points are relevant in the context of the following discussion where mathematics is seen to have limited success in limited domains in solving limited types of problems. Not only do we choose what to view, we also select, or rather create, the types of mathematics to use while ignoring those situations where mathematics simply does not work. Perhaps we reduce until we find some degree of success which is applauded without recourse to that which remains excluded.

3. A systemic functional approach to multisemiosis in mathematics

Mathematics is viewed as multi-semiotic construction where language, the visual images and the symbolism each possess specialised grammars designed to fulfill specific functions. Following Halliday (1994), meaning is made through the available choices within the grammatical systems which are organised metafunctionally. There are different systems for each semiotic resource for interpersonal meaning (engagement with the reader), experiential meaning (construction of experience), logical meaning (establishment of logical relations) and textual meaning (organisation of the text). Together these four categories of abstract systems constitute the grammar of each semiotic resource.

Some examples of grammatical systems in symbolic mathematics are given here for illustrative purposes only. If we consider the statement $y = 2x^3$, in terms of *experiential meaning*, the entity is identified as being equal to $2x^3$ in what forms a Token/Value relationship through the Relational process " = ". However, rather than being frozen in space and time, the mathematical statement involving y and $2x^3$ is a dynamic complex involving the mathematical participants (y, 2 and x), and the processes of identification (=) and multiplication ($2 \times x \times x \times x$). If this relation is displayed visually on a computer screen, it unfolds spatially in nanoseconds of time. In terms of *interpersonal meaning*, the choice from the system of modality (which acts to convince the reader of the truth value of the statement) is realised through the selection of " = ". This gives an absolute value of probability rather than a lowered value which could be realised through the symbolic equivalent of "might = ", "could = " or "would = ". Uncertainty in mathematics is

typically expressed through other means, for example probability statements. *Logical meaning* is realised through linguistic and symbolic conjunctive adjuncts and structural conjunctions aided by textual layout. In terms of *textual meaning*, the statement is organised so that the *y* appears on the left hand side of the equation. Symbolic mathematics is organised in very specific ways to make immediate the experiential, logical and interpersonal meaning of the mathematical statements.

There are multiple metafunctionally based systems which constitute the different grammars of the symbolism, visual display and visual images. Critically, choices from these systems function not only *intra-semiotically* within their own grammars, but they can also be re-represented *inter-semiotically* across the three different grammars with, in some cases, an unprecedented degree of equivalence of meaning. For example, the linguistically realised item "distance" functions intra-semiotically within the grammar of the English language. However, "distance" can also be inter-semiotically re-represented in the form of a visual line segment and a symbolic *x*. Thus (i) "distance", (ii) "——" and (iii) "*x*" each function as choices in three different grammars. However, once these inter-semiotic relations are established across language, visual images and the symbolism, the different choices become somewhat interchangeable. For example, the symbolic may appear in the linguistic and visual parts of the text. The most versatile semiotic resource appears to be the symbolism where functional elements sit comfortably not only within symbolic statements, but also within linguistic and visual forms of semiosis. The sophisticated inter-semiotic relations in mathematics, however, do not always result in equivalence or congruency resulting in an interchangeability of functional elements. As we shall see below, inter-semiosis opens up the way for semiotic metaphor where there is semantic change. Understanding this process may be one important key to understanding the success of mathematics.

The use of multiple semiotic resources means that during the process of inter-semiosis *semiotic metaphors* involving a semantic shift may occur as elements are translated into another semiotic form (O'Halloran 1999a, 1999b, 2003a, 2004). For example, a linguistically realised process such as "measuring" may become a visual entity in the form of a line segment in a mathematics diagram. This is a semiotic metaphor where a (linguistic) process becomes a (visual) entity in the shift from language to visual means of representation. This means that the meaning potential of the visual realm can now be exploited in what is essentially a re-casting of the semantic realm. The mental/behavioural/material process of measuring becomes a visual entity which may now be re-conceptualised in relation to other visual entities in the mathematical diagram. The significance of such a semantic re-organisation becomes clear in Section 4, which is concerned with the nature of changes in visual representations in mathematics which occurred with Descartes.

At this stage, we may conjecture that the degree of metaphorical transference across semiotic resources exceeds that possible within one semiotic resource such as language. It appears that the combined use of the symbolism, visual display and language is the critical point with regard to the success of mathematics in creating meaning which extends beyond our commonsense understanding of the world.

Mathematics is therefore seen as a tool that orders the world in particular ways according to the meaning potential arising from system choices within and across the three semiotic resources. This includes a heightened propensity for metaphorical transference through semiotic metaphor. As I have argued elsewhere (O'Halloran 2003b), this metaphoricity across different forms of semiosis may have contributed to the development of grammatical metaphor in language.

The aim of Section 4 is to develop an understanding of what mathematics was designed to do and the means through which this was achieved. A historical examination of the beginnings of modern mathematics and science as displayed in visual images and the work of Descartes and Newton is undertaken. We may see that the nature of the theological, intellectual and technical project of these times led to the creation of a mathematical and scientific reality which was grounded in a well defined semantic domain. Mathematics is perhaps effective in science because it provides the semiotic tools upon which the enterprise was premised.

4. A historical perspective of mathematics

In the early Renaissance, we see visual semiotic constructions in mathematics where the human body, sensual experience and circumstantial features provide the context for the formulation of mathematical concepts, ideas and problems. For example, Gregor Reisch's (1496) formulation of the concept of a solid in geometry is displayed in Figure 1.

A solid is explained as having length, depth and breath through the display of spears penetrating the figure of a man firmly implanted on a small mound on the ground. The penetration of spears through the human body was presumably a socially acceptable and respectable way of conceptualising the notion of a solid. To determine if these types of representations were a reaction to the dry abstraction of earlier texts such as Euclid's Elements (or indeed whether earlier Greek mathematics contained such types of drawings) deserves further investigation.

The physical context of the mathematical problem was also often portrayed visually during the early Renaissance. For example, a man attempts to calculate the inner dimensions of what appears to be a castle well or tower in Figure 2.

Similarly, Tartaglia's drawing depicts a man calculating the height of cliff upon which a castle is located in Figure 3.

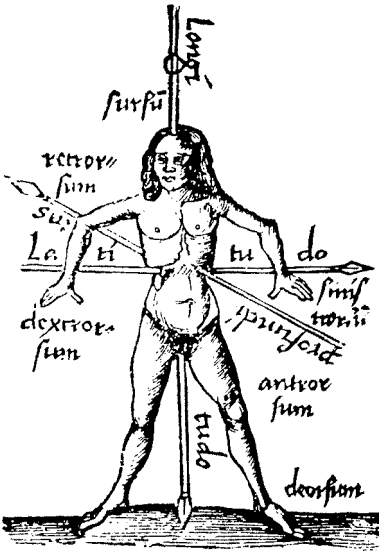


Figure 1. Concept of volume in Gregor Reisch's (1496) *Margarita Philosophica* (1535 reprint). Reproduced by courtesy of the Director and Librarian, the John Rylands University Library of Manchester

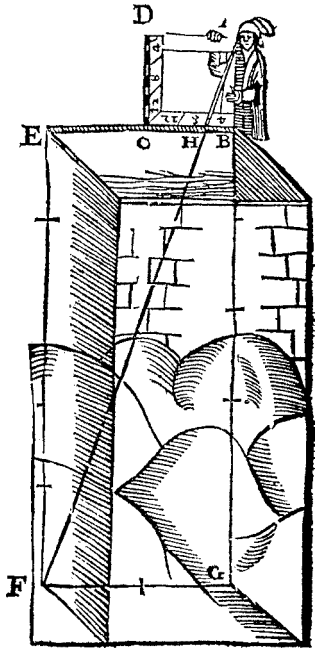


Figure 2. Calculating depth and width of a Castle Well or Tower Illustration (Babington 1635). Reproduced by permission of the Syndics of Cambridge University Library

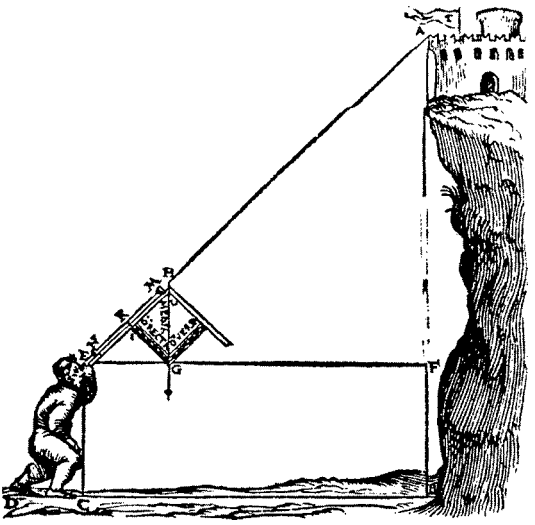


Figure 3. Calculating the Height of the Cliff Face (Tartaglia 1546). Reproduced by courtesy of the Director and Librarian, the John Rylands University Library of Manchester

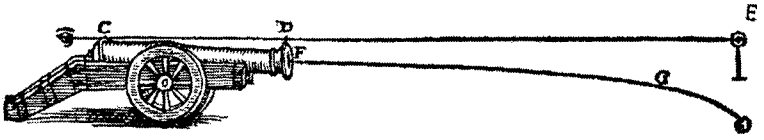


Figure 4. The Path of the Cannon Ball (Tartaglia 1546). Reproduced by courtesy of the Director and Librarian, the John Rylands University Library of Manchester

As shown in Figure 4, Tartaglia was directly concerned with the calculation of the path of missiles (see Davis 2000; Høyrup 1994; O'Halloran 2003a). Other visual representations in mathematics include problems such as calculating the distance across a river as depicted, for example, in Belli (1565). In one of Belli's illustrations (see Eagle 1995: 52) the lines of sight of the men are drawn against a backdrop of the river and surrounding countryside.

What is extraordinarily interesting about these early mathematical drawings, such as those shown in Figures 1–4, is that a commonsense material world and the objectives of human activity within that world are clearly depicted. The mathematics is functional in formulating and solving particular physical problems which are explicitly depicted. The images contain human figures who are involved in some type of physical or mental activity. The material context of the mathematics problem is displayed: how high is a castle wall, how to gain accuracy from cannon fire, and what distance across the river. Although stylistically different, these types of mathematical drawings still appear in texts such as high school and university

textbooks. However, modern illustrations either serve to introduce problems in order to contextualise the mathematical theory which is to be presented, or they are exercises to practise the application of the theory. However, the images in early Renaissance texts appear as precursors to the development of theory itself. It is precisely this environment that modern mathematical textbooks attempt to emulate before moving into contemporary mathematical theory.

In visual constructions of the context for the mathematical problems, new semiotic entities are created; ones that do not exist in our visual perceptual world. That is, lines are drawn for distances, heights and lines of sight, and curves are drawn for the path of the cannon ball and missiles. This is viewed as a case of semiotic metaphor where the use of a semiotic resource, in this case, the visual semiotic, allows for the introduction of new metaphorical entities. That is, entities such as lines of sight do not exist as material participants or objects, but rather only exist as semiotic construals of that world. In what follows, it becomes obvious that these semiotic lines and curves became the central area of concern in mathematics, and this occurred for a very good reason. The new metaphorical entities led to the construction of further metaphorical visual entities which could be described symbolically as a series of relations. This represented a significant turning point in the development of mathematics.

The nature of mathematical drawings changed over time, especially with the work of Descartes and Newton. Lines and curves tended to become more promi-

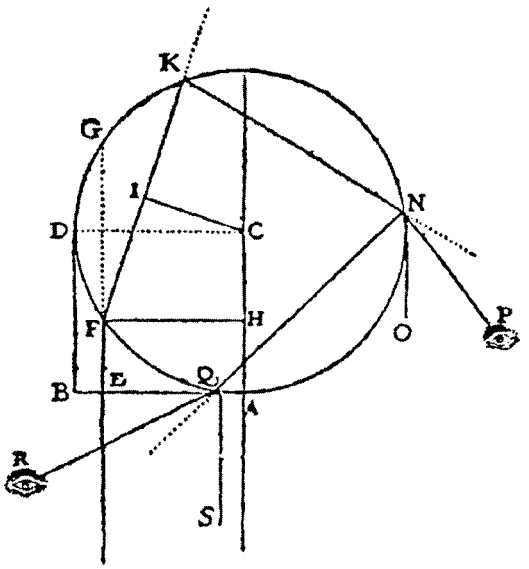
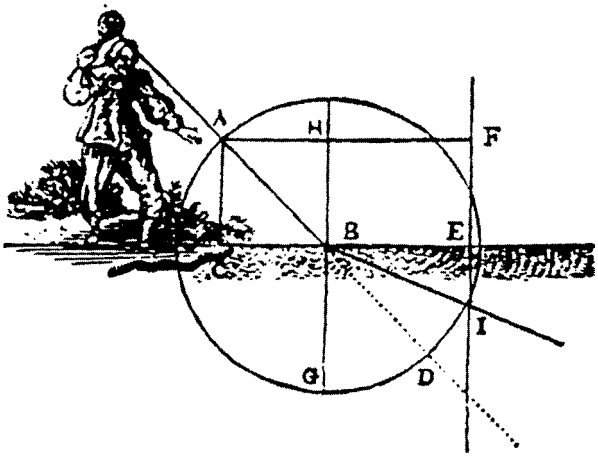


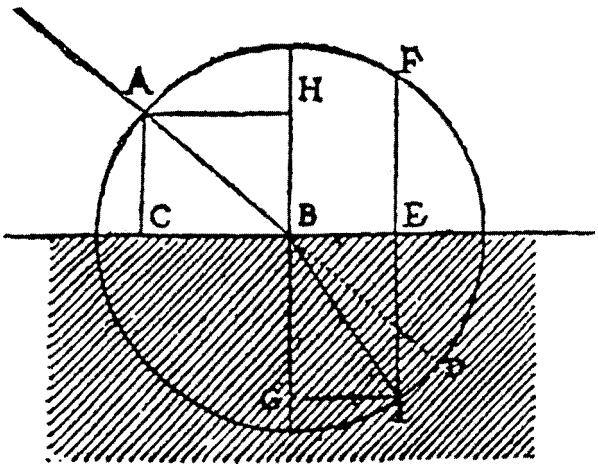
Figure 5. The Human Eye Remains (Descartes 1998: 92). Reproduced by courtesy of Cambridge University Press

nent while participants such as the men, and circumstances like cliffs, rivers, castles and cannons increasingly disappeared. In some cases, there only remained one part of the human body, the eye as displayed in Figure 5.

There was a shift in the nature of the depicted activity in this new type of semiotic reconstrual. Men were no longer engaged in some physical, mental or material activity such as measuring, but rather the human eye minus the body became engaged in acts of perception.



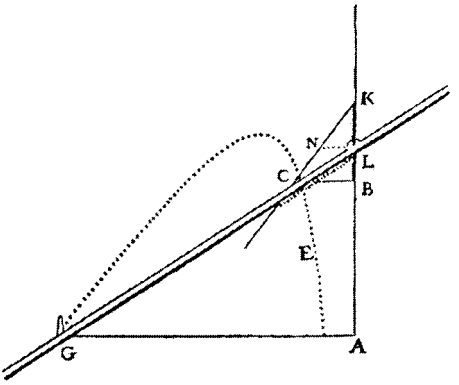
A



B

Figure 6. From Context to Circles and Lines. (A) Descartes (1998: 79), (B) Descartes (1998: 81). Reproduced by courtesy of Cambridge University Press

LIVRE SECOND. 321



Aprés cela prenant vn point a diſcretion dans la courbe, comme C, ſur lequel ie ſuppoſe que l'inſtrument qui ſert a la deſcrire eſt appliqué, ie tire de ce point C la ligne C B parallele a G A, & pource que C B & B A ſont deux quantités indeterminées & inconnuës, ie les nomme l'vne y & l'autre x. mais affin de trouuer le rapport de l'vne à l'autre; ie conſidere auſſy les quantités connuës qui determinent la deſcription de cete ligne courbe, comme G A que ie nomme a, K L que ie nomme b, & N L parallele à G A que ie nomme c. puis ie dis, comme N L eſt à L K, ou c à b, ainſi C B, ou y, eſt à B K, qui eſt par conſequent $\frac{b}{c}y$: & B L eſt $\frac{b}{c}y - b$, & A L eſt $x + \frac{b}{c}y - b$. de plus comme C B eſt à L B, ou y à $\frac{b}{c}y - b$, ainſi a, ou G A, eſt á L A, ou $x + \frac{b}{c}y - b$. de façon que multipliant

Sf

Figure 7A. Descartes' Use of Algebra to Construct Curves. Descartes (1954:53). Reproduced with acknowledgement and thanks to Dover Publications

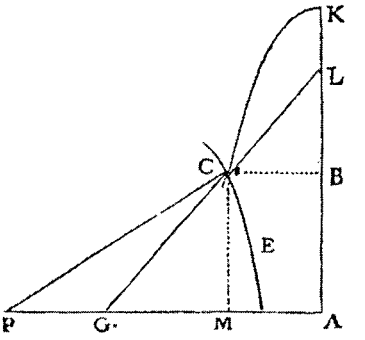
In Figure 6, the removal of the physical context and the human body is clearly demonstrated in the beginnings of modern mathematical formulations.

In Figure 6A, Descartes' drawing of the path of a ball travelling through a finely woven cloth (indicated by the line segments AB to BD) and water (AB to BI) includes the man actually hitting the ball. In Figure 6B, however, the physical context is removed as the path of the ball is depicted as a series of lines and a circle.

LIVRE SECOND. 349

Puis par le troiſieſme terme il faut chercher g la ſeconde quantité, & on a $gg \infty 3ee - 4be - 2cd + bb + dd$. Puis par le penultieſme il faut chercher h la penultieſme quantité, qui eſt $h^3 \infty \frac{2bbccd\acute{d}}{e^3} - \frac{2bccdd.}{ee}$ Et ainſi il faudroit continuer ſuiuant ce meſme ordre iuſques a la derniere, s'il y en auoit d'auantage en cete ſomme; car c'eſt choſe qu'on peut touſiours faire en meſme façon.

Puis par le terme qui ſuit en ce meſme ordre, qui eſt icy le quatrieſme, il faut chercher la quantité v, & on a



$$v \infty \frac{2e^3}{dd} - \frac{3bee}{dd} + \frac{bbe}{dd} - \frac{2ce}{d} + e + \frac{2bc}{d} + \frac{bcc}{ee} - \frac{bbcc}{e^3}$$

ou mettant y au lieu d'e qui luy eſt eſgal on a

$$v \infty \frac{2y^3}{dd} - \frac{3byy}{dd} + \frac{bby}{dd} - \frac{2cy.}{d} + y + \frac{2bc}{d} + \frac{bcc}{yy} - \frac{bbcc.}{y^3}$$

pour la ligne A P.

Et ainſi la troiſieſme equation, qui eſt

Xx 3 &c.

Figure 7B. Descartes' Use of Algebra to Construct Curves. Descartes (1954: 109). Reproduced with acknowledgement and thanks to Dover Publications

This example serves as an introduction to Descartes' discussion of the refraction of light which is visualised in a manner similar to Figure 6B.

The removal of human activities of action and eventually perception in the new mathematical drawings meant that the nature of the depicted processes became in a sense spatial and relational. In other words, if we consider the major

participants to be the circle and line segments, the concern lies in the relations between those entities and their parts rather than the material context of the problem. Replacing the semiotic construal of the material with the metaphorical in the form of mathematical diagrams involving lines, triangles and arcs permitted the solution of new types of problems. The height of the cliff and the castle wall were at least theoretically physically measurable. But how so with the path of a cannon ball? This became possible only through a concern with spatiality and relations which Descartes eventually linked to algebra description. In time the visual images assumed a secondary status in relation to the algebra (see e.g. Davis 1974). Descartes' construction of different curves and an increasing reliance on algebra to describe those curves may be seen in Figures 7A and 7B.

Descartes had shifted from the construction of curves using the Greek material compass and ruler to using an abstract compass which was semiotically grounded. Descartes claimed that his proportional compass had the same certainty as the ordinary compass. As Shea (1991:45) explains: "This new instrument does not have to be physically applied; it is enough to be able to visualise it and use it as a computing device. In other words, pen and paper is all that is required, since the nature of the curve is revealed in its tracing."

The beginnings of modern mathematics are seen in Descartes where the curves are described algebraically although, as Davis and Hersh (1986) claim, this really amounted to algebraisation of ruler-and-compass constructions. "In its current form, Cartesian geometry[2] is due as much to Descartes' own contemporaries and successors as to himself" (Davis & Hersh 1986:5). Although later mathematicians considered the geometrical construction to be sets of points satisfying certain criteria defined by the algebraic equation, despite an increasing dependence on algebra, "Descartes never *defined* as geometrical those curves that admit of algebraic equations" ... [nonetheless] ... He simplified algebraic notation and set geometry on a new course by his discovery that algebraic equations were useful not only in classifying geometrical curves, but in actually devising the simplest possible construction" (Shea 1991:67).

Mathematicians such as Newton later used the algebraic equations as complete descriptions of curves rather than a tool for construction. Mathematical symbolism became the semiotic through which curves were defined and problems solved with the aid of mathematical graphs and diagrams. However, we may ask what prompted Descartes to algebraicise his geometry, thus paving the way for "his new world of relations that seventeenth century mathematicians entered with pride" (Shea 1991:67). We must turn to Descartes' philosophy in order to understand how these mathematical construals, which had became decontextualised and algebraicised, represented a method upon which he thought was the path to knowledge. He used algebra as a tool for the construction of curves, but perhaps

from this endeavour arose his method for construing and deriving what he viewed as true. And that method depended on algebraic descriptions which, while offering more, admitted less.

5. The Cartesian project

Descartes' project was to solve the problem of knowing what is true at the cost of denying what appears to be obvious. Following Plato (429–347BC), Descartes turned to reason rather than the senses as the means for achieving that aim. For Descartes, sensory perception of the material world was unreliable in a way similar to that envisaged by Plato. While Plato used the narrative of shadows in a cave to illustrate sensory illusion, Descartes attempted to demonstrate the unreliability of the senses through a discussion of a ball of wax in the Second Meditation (1952: 202–212). Descartes explains that the properties of wax perceived by the senses, for example, flavour, smell, colour, shape and size, are unreliable because they change as the wax is heated. Mental perception rather than sense perception allows examination of the reliable essence of matter which he conceptualises as motion and extension.[3]

> What, then, was it I comprehended so distinctly in knowing the piece of wax? Certainly, it could be nothing of all that I was aware of by way of the senses, since all the things that came by way of taste, smell, sight, touch and hearing, are changed, and the wax none the less remains... As we find, what then alone remains is a something extended, flexible and moveable... which cannot be [adequately] apprehended save by the mind. (Descartes 1952: 208–209)

Descartes' method is described in the twenty one *Rules for the Direction of the Mind*, which was later reduced to a method involving evidence, division, order and exhaustion in *Discourse on Method*, (Part II). Descartes' method was to divide to get the simplest essence, and then order and enumerate to understand composites. As Shea (1991: 131) explains: "(a) nothing is to be assented to unless evidently known to be true; (b) every subject-matter is to be divided into the smallest possible parts, and each dealt with separately; (c) each part is to be considered in the right order, the simplest first; and (d) no part is to be omitted in reviewing the whole."

Descartes' method reveals his increasing dependence on his new semiotic tools in the form of algebraic descriptions. In other words, it appears that his four cannons or rules were built on what he could achieve semiotically through his algebraicisation of geometry. He could express relations in the simplest elements, and then rework to understand the more complex through the use of the symbolism. His dissatisfaction with the inadequacies of language is openly expressed

in the Second Meditation in *Meditations on First Philosophy* when he attempts to describe what is knowable about the ball of wax:

> I am indeed amazed when I consider how weak my mind is and how prone to error. For although I can, dispensing with words, [directly] apprehend all this in myself, none the less worlds have a hampering hold upon me, and the accepted usages of ordinary speech tend to mislead me. . .. But aiming as I do at knowledge superior to the common, I should be ashamed to draw grounds for doubt from the forms and terms of ordinary speech. (Descartes 1952: 209–210)

Descartes wanted an aid for thought for describing the essence of bodies, that which is knowable in terms of motion and extension, and this aid should be simple and abstracted from anything superfluous. For "perfect understanding", the question should be rendered "as simple as possible, and resorting to enumeration, divide[d] . . . into its minimal parts" (Descartes 1952: 76). In addition to using the geometrical curves as an aid to thought,[4] Descartes explicitly says in Rule XVI that mathematical representations should be algebraically expressed due to the simplicity of this type of formulation.

> Thus if I write $2a^3$, that will be as if I should write the double of the magnitude signified by the letter a, which contains three relations. By this device not only do we obtain a great economy in words, but also, what is more important, we present the terms of the difficulty so plain and unencumbered that, while omitting nothing which is needed, there is also nothing superfluous, nothing which engages our mental powers to no purpose. . . (Descartes 1952: 101)

It is not clear how Descartes developed his method of algebraically describing curves. According to Shea (1991: 48–49), "We do not have his account of the evolution of his ideas on this point, but one crucial step was surely the discovery that expressions such as "square" for x^2 or "cube" for x^3 do not necessarily represent a given shape". In Rule XVI Descartes explicitly says that terms such as the square, the cube and the biquadratic had long misled him and should be abandoned (Descartes 1952: 101–102).

Descartes perhaps recognised the importance of representing "the cube" as $2a^3$ because "the cube" is a nominal group which semantically is a fixed entity. On the other hand, $2a^3$ is a complex of participants and processes which may enter into a relation such as $y = 2x^3$. This relation may be rearranged and also differentiated from other "cubes" such as $y = 3x^3$ which has a different shape. The linguistic "cube" is not equivalent to the symbolic formulation. Descartes' self professed confusion concerning the linguistic construction demonstrates just how difficult it is to escape from traditional semiotic frameworks in order to move into new ways of thinking, perceiving, being and creating.

In trying to find the means to construct curves based on proportionality, Descartes described that proportionality as simply as he could, and that involved replacing line segments with abstract quantities. He described algebraically the nature of the proportional relations between two quantities as a tool to aid construction and classification of curves. Descartes' preoccupation to draw these shapes meant that he failed to see that these equations were complete descriptions or representations of the curve. For Descartes, although equations incorporate information about the properties of curves, they do not provide sufficient representation of their geometrical reality.

> The next generation of mathematicians were satisfied with the equations, and did not concern itself with actually forming the construction of curves.
>
> (Shea 1991:60)

For Descartes, the advantages in using algebraic descriptions for curves were numerous. Algebraic formulations were representations of relations consisting of the simplest elements. They had a direct relation to the curves in terms of the proportionality which was displayed graphically, and the formulations could be used as tools for reasoning. The result of these efforts was that mathematics became concerned with a limited semantic field in the form of proportional relations which could be displayed visually. The algebraic expression of those relations were concerned with brevity and the removal of extraneous information for effective reasoning. This resulted in the development of a symbolic grammar which condensed meaning most simply and efficiently in the form of the essential elements. As I discuss in the final section of this chapter, the resulting grammar for mathematical symbolism developed a new type of grammatical complexity which is not found in language.

6. Newton's view of the world

Descartes defines motion, size, shape and arrangement of parts as the concern of mathematics in an attempt to remove doubt through the use of the intellect. The instrument through which certain conclusions should be reached involved semiotic construals of that phenomenon in the form of algebraic descriptions and curves. This led to a certain homogenisation where material objects became essentially identical and replaceable with one another. In addition, the material context was removed from that semiotic space. The move from perception and natural appearance is apparent in the semiotic nature of mathematical texts. The human body and eye were also removed. The material world developed a secondary status as it was replaced with what was mathematically describable in terms of spatiality and relations through symbolism and visual images. Descartes' project entailed a

re-writing of nature, as explained by Barry (1996: 55): "From the Cartesian point of view, the only proper way to decode nature is to write it ... in the most exact terms possible." But from the preceding discussion, we can see that this description centred around particular dimensions of meaning.

The modern mathematical and scientific project did not end with Descartes. Instead Isaac Newton developed a science which included rather than excluded matter. Rather than depending on Cartesian notions of extension and motion, Newton introduced physical matter in order to formulate entities such as attraction, force, absolute space and time, and gravitation. As Barry (1996: 55) explains, Newton altered the Cartesian project of the primacy of ideas and the dichotomy of mind and matter in relation to that which was mathematically intelligible. Newton's semiotic is mathematical symbolism (see Figure 8) and he conceptualises the invisible such as forces of attraction and gravity as mathematical laws.

In this way, Newtonian attraction is "written in a syntax inspired directly by the mathematical forms which Descartes (and Galileo) stressed time and time again as the most proper form of articulation for the new physics" (Barry 1996: 127). We may not be able to perceive attraction, but semiotically it can be construed. "The decisive importance of the Newtonian mathematical vision regarding the issues of natural appearance and perception is dramatically revealed in the concept of attraction: what we see with our eyes demands the existence of what may never be seen, but must be mathematically granted" (Barry 1996: 134). The division between the perceptually visible and invisible disappears with Newton.

Newton's theoretical stance is accompanied by real and imaginary experiments to support or refute his theories. In this shift from the real to the experimental, once again what is semiotically constructed is accepted as scientific validation.

> Newton believes that many of the experiments which he devised but never fully implemented (e.g. his notebook drawings) nonetheless serve as empirical validation or invalidation of certain hypothetical stances. It is as if the diagram of a possible experiment is virtually the same as the experiment itself.
>
> (Barry 1996: 160)

Descartes sought to move beyond the senses and realm of appearance to the intellect to describe mathematically that which could not be doubted in the form of shape and motion. Newton enlarged this project by admitting matter and that which could not be perceived to be construed mathematically. This project was aided by the use of technical and laboratory apparatus where the phenomena under examination were semiotic abstractions. While recent developments in mathematics in the past two decades have fundamentally changed the Newtonian mechanical view of the universe (for example, dynamical systems theory is concerned with the nonlinearity rather than linearity of physical systems), mathematical symbolism, visual display and language remain the major resources through

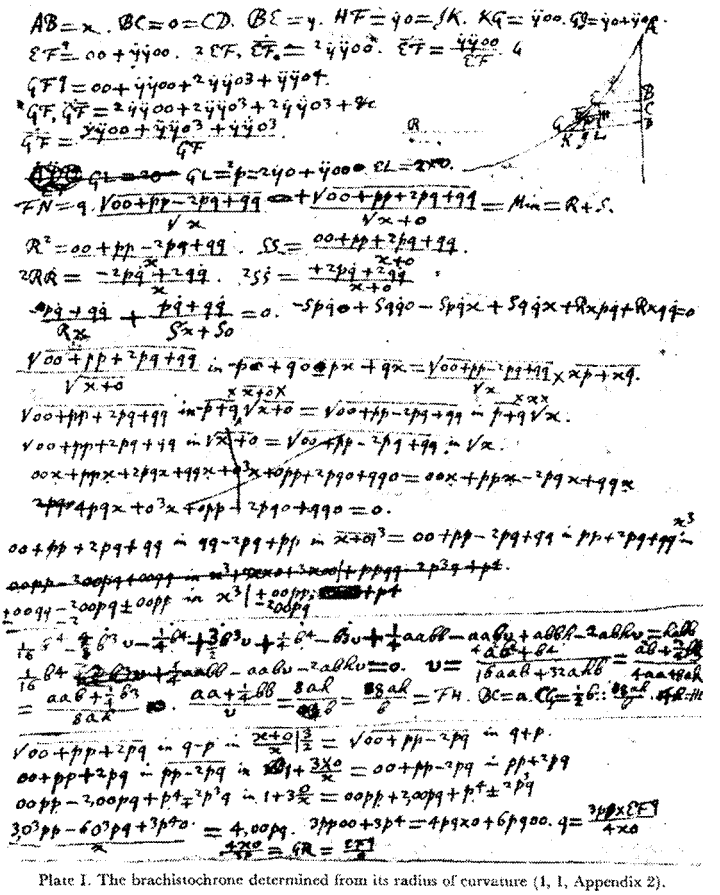


Plate I. The brachistochrone determined from its radius of curvature (1, 1, Appendix 2).

Figure 8. Newton's Use of Algebra reproduced from D. T. Whiteside's *Mathematical Papers of Isaac Newton* (Newton 1981: opposite p. 90). Reproduced by courtesy of Cambridge University Press

which physical systems are described, although now this is largely accomplished computationally in a dynamic virtual and often visual world through the use of computers. Nature was re-written semiotically, and in Section 7, I briefly examine the nature of the semantic realm which was admitted in this re-construal. For Newton and others "Mathematical precision is the only form which the perfect relation between mind and matter could possibly take" (Barry 1996: 127). What was allowed within that field of mathematical precision?

7. Meaning in modern mathematics

Following Halliday's (1994) metafunctional hypothesis of language which is extended to mathematical symbolism and visual display, I discuss in general terms the nature of experiential, logical, textual and interpersonal functions of language, visual display and mathematical symbolism typically found in mathematical discourse. To aid this discussion, I use a contemporary formulation of volume (Stewart 1999: 380) which may be compared to the earlier conceptualisation of volume displayed in Figure 1. The final part of this discussion is concerned with intersemiosis where the interaction between the resources becomes the critical factor in the success of mathematics to make meanings beyond that achieved via our linguistic and pictorial repertoires. From this point I return to the question of the success of mathematics in the physical sciences.

a. Language

Show that the volume of a sphere of radius r is $V = \frac{4}{3}\pi r^3$.

Experiential Meaning: Language functions to provide the context for mathematical concepts, definitions, problems, solutions and the implications of the results. The mathematics linguistic text tends to be lexically dense with a high incidence of grammatical metaphor (for example, nominalisation where processes are realised as entities; e.g. intersect becomes "intersection"). There are also extended nominal groups, for example "the volume of a sphere of radius r". This makes for a dense and static type of text which is typical of mathematical and scientific writing (Halliday & Martin 1993).

Logical Meaning: The procedures are explained as a series of steps and the implications of the findings are formulated and contextualised through language.

Textual Meaning: The organisation of the linguistic text (and the whole mathematics text) is generically defined.

Interpersonal Meaning: The linguistic text is usually a series of statements and commands depending upon the genre of the mathematics text (for example, definition, explanation of theory and the solution to problems). In the above example of the volume of a sphere, the reader is commanded through an imperative to "Show that the volume of a sphere of radius r is $V = \frac{4}{3}\pi r^3$". The absence of modulation means maximal obligation for the reader. The objective rational view of mathematics is achieved in part through the lack of modalisation of statements,

in this case "the volume of a sphere of radius r is $V = \frac{4}{3}\pi r^3$". In keeping with Descartes, the apparently deceptive realm of the human is cast aside to form a discourse of certainty.

b. Visual images: Graphs and diagrams

Experiential Meaning: The relations displayed in mathematical visual images tend to match our perceptions of the material world. However, we move from the material to an idealisation which features new participants which formerly did not exist. For example, the x and y axis, the origin, the radius r and the triangle are introduced in Figure 9. The visual grammar is sophisticated with special systems of meaning (for example, the radius of the sphere is formulated in relation to x and y axis). The perceptual relations between those entities as relations of parts to the whole and patterns of continuity are made visible in readiness for the symbolic descriptions. The decontextualised visual image typically includes only that which is necessary for the mathematical solution to the problem.

Logical Meaning: The visual image allows the logical relations of the parts and patterns to be perceived and conceptualised but not described exactly.[5]

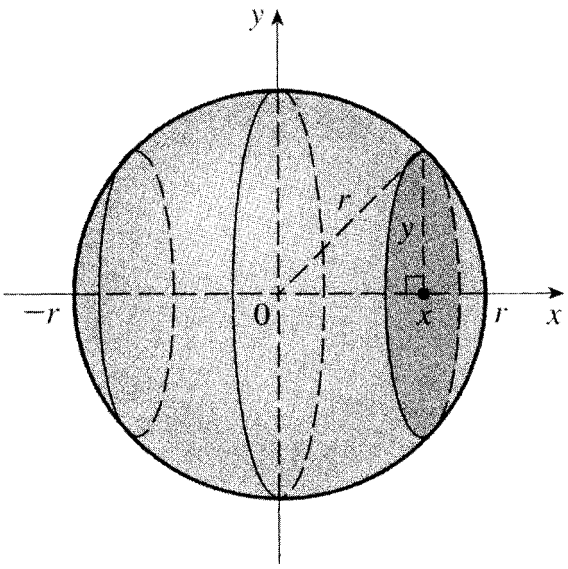


Figure 9. The Diagram for the Volume of a Sphere. Reproduced from Stewart (1999:380). From Calculus: Combined Single and Multivariable 4th edition by Stewart. © 1999. Reprinted with permission of Brooks/Cole, a division of Thomson Learning: www.thomsonrights.com

Textual Meaning: The visual images organise the proportional relations between the entities as a whole. The relative positioning indicates stability in the case of axis, and dynamicism in case of the curves. The labelling of the mathematics participants allows cohesive links to the main body of the text and the symbolism.

Interpersonal Meaning: The decontextualised nature of visual images in mathematics assumes a high modality. The images appear as perfect abstractions where the disorder, deception and ambiguity of material world have been removed.

c. Mathematical symbolism

$$V = \frac{4}{3}\pi r^3$$

This symbolic statement may be considered as a series of rankshifted participants ($\frac{4}{3}$, π and r) and processes ($\div$ and $\times$) as indicated by the square brackets [[ ]] below.

$$V = (4 \div 3) \times \pi \times r \times r \times r$$

$$V = \Big[\Big[\big[\big[[(4 \div 3)]\big] \times \pi\big]\Big] \times \Big[\big[\big[[r \times r]\big] \times r\big]\Big]\Big]$$

Experiential Meaning: The exact relations between the mathematical participants are captured symbolically. The grammar of mathematical symbolism functions to construe relations dynamically so that these configurations may be rearranged in progressive steps for the solution to the mathematics problem. For example, the formula may be reconfigured so that r becomes the subject rather than V. The symbolism developed as the tool for the solution to the problems, and the grammar thus takes a specific form to fulfill this role. For example, there is limited range of process types. The major processes are Relational (identifying equalities and inequalities and so forth) and Operative (mathematical processes of addition, subtraction and so forth). The grammar has systems such that there is economy and precision without repetition. Relations are expressed concisely as possible, but in a matter which allows reconfiguration. New systems are introduced in the grammar of the symbolism that do not exist in language. For example, spatial positioning has meaning (for example, x^2 and x_2) and brackets are used to change the conventional rule of order of the mathematical processes which differs from the left to right syntagmatic organisation of language. Rather than condensing meaning in nominal group structures, the strategy through which the grammar of mathematical symbolism operates is *rankshifting* the configurations of mathematical processes and participants. This combination of grammatical intricacy and gram-

matical density results in grammatical complexity which is not found in language. The human agent rarely features other than to follow particular instructions.

Logical Meaning: The symbolism is the semiotic tool through which logical reasoning largely takes place. In the solution to mathematics problems, long sequences of symbolic complex nested structures of reasoning take place.

Textual Meaning: The textual organisation of the symbolism has generic conventions which allow the solution to the mathematics problems to be followed with comparative ease. There are discursive links to the main body of the text through linguistic selections.

Interpersonal Meaning: The symbolic statements and commands lack modality and hence have a certain sense of absoluteness in terms of truth. The style of production of mathematical texts, and the abstract nature of the participants also lend an idealism of truth. There is a lack of that which approximates the finite human world.

8. Inter-semiosis in mathematics

Meaning is made in mathematics intra-semiotically through the grammars of language, visual images and mathematical symbolism and inter-semiotically across the three semiotic resources which results in semantic equivalence or semiotic metaphor. The success of mathematics lies in the interlocking of systems of meaning across the symbolism, visual display and language.

In the case of the volume of a sphere, the *r* means through semantic equivalence. That is, the *r* can mean linguistically as the abstract mathematical radius which stands for the specific through the general, and so the meaning potential of language is exploited. The *r* can also mean visually through the grammar of mathematical visual semiosis as seen in the diagram of the sphere. The meaning potential of the visual semiotic is also utilised. In the diagram, the relations between the participants can be perceptually understood. The *r* means symbolically as a participant which enters into specific relations with other participants. This relation cannot only be completely described, but also changed in order to solve problems.

Each time there is a shift between the semiotic resources, a new meaning potential is exploited. Also metaphorical constructions can occur. For example, "the volume of a sphere of radius *r*" is an entity, a nominal group in language. However, visually "volume" becomes a set of spatial relations between the *x* and *y* axes

and r. When described symbolically, "volume" is defined as a complex of relations involving mathematical participants $\frac{4}{3}$, π and r and the processes of division and multiplication. The metaphorical linguistic version of volume is unpacked symbolically, and this indicates that there exists some form of relationship between semiotic and grammatical metaphors (O'Halloran 2003a). Intersemiosis permits metaphorical transferences of meaning together with the opportunity to use the systems of meaning within the new semiotic.

Descartes' desire to construct and differentiate between curves resulted in the development of symbolic descriptions. His algebra linked real world problems (the path of light) to visual realisation (the lines, line segments and circles) to complete description (symbolic statements). Proportionality was best described symbolically because, as he himself admitted, words were misleading as they did not possess the meaning potential to differentiate exactly between curves. The symbolism, however, developed such a potential. Descartes thought it was the construction of the curve that was critical, but those following focused on the symbolism. The remaining question is why are mathematical proportional relations expressed symbolically so successful in describing the physical world? One answer which has already been suggested is that the mathematical/scientific order which has been imposed on our conception of the world has limits to its success.

9. Complexity, chaos and Wigner on the effectiveness of mathematics

Since Descartes, our understanding of the ordered physical world has changed in that the complexity, chaos and indeterminacy underlying physical systems is now generally accepted. Despite this, the step by step approach of modern science has been reasonably successful in describing the behaviour of physical systems because, as Davies (1990) explains, many systems are approximately linear in nature and so breaking down the systems into smaller parts to understand the nature of the whole as advised by Descartes appears to work: "By analysis, one can chop up complicated systems into simpler components. And understanding of the behaviour of the components then implies, *ipso facto*, an understanding of the whole" (Davies 1990). However, this attempt to define mathematically the underlying regularity of ordered systems breaks down at some point when the behaviour of the system becomes unpredictable:

> On the other hand, they [all physical systems] turn out to be nonlinear at some level. When nonlinearity becomes important, it is no longer possible to proceed by analysis, because the whole is now greater than the sum of the parts. Nonlinear systems can display a rich and complex repertoire of behaviour and do unexpected things.

> Generally speaking, a nonlinear system must be understood in its totality, which in practice means taking into account a variety of constraints, boundary conditions and initial conditions. (Davies 1990: 16)

The unpredictable behaviour of chaotic dynamical systems can only be charted and studied with the new semiotic tools afforded by technology. Pen and paper mathematics has been replaced by computer generated algorithms and virtual worlds.[6] The realm of the technical thus plays an important role in defining the nature of semiosis, and consequently the nature of meanings that can be made. Computers mean, for example, that dynamic visual images are now possible, and colour has become a system that may be utilised in mathematical descriptions of fractals and dynamical system problems. Today, visualisation plays a great role through the use of computers and other technologies through which mathematical models of the world are semiotically constructed.

So is Wigner's question of the reasonable effectiveness of mathematics a straw house? After all, the success of mathematics in the sciences may be seen to be due to the fact that: (1) following Descartes, we see what we decide to look for; (2) the success of that search depends on the available semiotic tools; (3) those tools consist of language, visual images and symbolism which function intra-semiotically and inter-semiotically; (4) our forms of semiosis in turn determine what is admitted into the scientific realm; (5) our semiotic tools evolve with available technology; (6) we discard the many cases in which mathematics does not work; and (7) contemporary mathematics shows the limitations of our scientific view of the world.

Regardless of the position adopted here, the removal of the material and the body in mathematics does not make its successes (or failures) mysterious.[7] As Hersh (1997) claims, mathematics is just another realm of human activity. At the cost of limiting the semantic field with which mathematics is concerned and dismissing the remainder as unreliable, particular types of descriptions, predictions and prescriptions have become possible and these up to a point have been successful in science. Perhaps that realm should be made more explicit in order to avoid applying mathematical tools to situations beyond that for which they were designed, as happens in social sciences, politics, economics and education, for example. And where mathematics can be used effectively, we could start appreciating more fully what has been cast aside in order to achieve that limited success.

Notes

* I wish to thank Emeritus Professor Philip Davis (Applied Mathematics Division, Brown University) for his valuable comments on an earlier draft of this chapter.

For further development of mathematics as a multisemiotic discourse see: O'Halloran, K. (2005). *Mathematical Discourse: Language, Symbolism and Visual Images.* London & New York: Continuum.

The illustrations in this chapter have been reproduced with the kind permission of the following libraries and publishers:

Figure 1: Reproduced by courtesy of the Director and Librarian, the John Rylands University Library of Manchester.
Figure 2: Reproduced by permission of the Syndics of Cambridge University Library.
Figure 3: Reproduced by courtesy of the Director and Librarian, the John Rylands University Library of Manchester.
Figure 4: Reproduced by courtesy of the Director and Librarian, the John Rylands University Library of Manchester.
Figure 5: Reproduced by permission of Cambridge University Press.
Figure 6A: Reproduced by permission of Cambridge University Press.
Figure 6B: Reproduced by permission of Cambridge University Press.
Figure 7A: Public domain material with acknowledgement and thanks to Dover Publications.
Figure 7B: Public domain material with acknowledgement and thanks to Dover Publications.
Figure 8: Reproduced by permission of Cambridge University Press.
Figure 9: Reproduced with permission from Thomson Learning Global Rights Group.

1. Advocates of a semiotically informed reading of mathematics include Rotman (1988, 2000), Lemke (1998) and Anderson et al. (2003).

2. As Davis and Hersh (1984:5) explain "Coordinate geometry, as currently taught, involves the placing of perpendicular axes in a plane, the assignment of two coordinates (or addresses) to each geometrical point, and the replacement of straight lines and curves by appropriate algebraic equations".

3. Descartes (1637/1952) defines extension as "whatever has length, breadth and depth, not inquiring whether it be truly body or merely space" in Rule 14 from *Rules for the Direction of Mind.*

4. Descartes (1637/1952) states that it is often helpful to draw [simplified] figures in Rule XV.

5. However, with computer graphics the importance of the visual image is increasing as dynamic displays become possible as predicted by Davis (1974).

6. See Philip Davis (1998) on the use of computer graphics and links between mathematics, war and entertainment.

7. See Judovitz's (2001:83–107) discussion of the conceptual construction and the rhetorical and metaphorical use of the body in Descartes' *Meditations.*

References

Anderson, Myrdene, Sáenz-Ludlow, Adalira, Zellweger, Shea, & Cifarelli, Victor (2003). *Educational Perspectives on Mathematics as Semiosis: From Thinking to Interpreting to Knowing*. Ottawa: Legas Publishing.

Babington, John (1635). *A Short Treatise of Geometrie*. London: Thomas Harper. Reprinted 1971. Amsterdam and New York: Da Capo Press.

Barry, James (1996). *Measures of Science: Theological and Technical Impulses in Early Modern Thought*. Illinois: Northwestern University Press.

Davies, Paul C. W. (1990). "Why is the world knowable." In R. E. Mickens (Ed.), *Mathematics and the Language of Nature* (pp. 14–54). Singapore: World Scientific.

Davis, Philip J. (1974). "Visual geometry, computer graphics and theorems of perceived type." *Proceedings of Symposia in Applied Mathematics, 20*, 113–127.

Davis, Philip J. (1998). "Mickey flies the Stealth." *SIAM News, 31* (3). http://www.siam.org/siamnews/04-98/mickey.htm

Davis, Philip J. (2000). *The Education of a Mathematician*. Natwick, MA: A. K. Peters Ltd.

Davis, Philip J. & Hersh, Reuben (1986). *Descartes' Dream: The World According to Mathematics*. New York: Harcourt Brace Jovanovich.

Descartes, Rene (1952). *Descartes' Philosophical Writings*. N. Kemp Smith (Ed. and Transl.). London: Macmillan & Co.

Descartes, Rene (1954). *The Geometry of Rene Descartes*. D. E. Smith & M. L. Latham (Transl.). 1st ed. 1637. New York: Dover.

Descartes, Rene (1998). *The World and Other Writings*. S. Gaukroger (Ed. and Transl.). Cambridge: Cambridge University Press.

Eagle, Ruth M. (1995). *Exploring Mathematics through History*. Cambridge: Cambridge University Press.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: Arnold.

Halliday, M. A. K. & Martin, James R. (1993). *Writing Science: Literacy and Discursive Power*. London: Falmer.

Hamming, Richard W. (1980). "The unreasonable effectiveness of mathematics." *American Mathematical Monthly, 87* (2), 81–90.

Hersh, Reuben (1997). *What is Mathematics, Really?* London: Jonathon Cape.

Høyrup, Jens (1994). *In Measure, Number, and Weight: Studies in Mathematics and Culture*. New York: State University of New York Press.

Judovitz, Dalia (2001). *The Culture of the Body: Genealogies of Modernity*. Ann Arbor: The University of Michigan Press.

Lemke, Jay L. (1998). "Multiplying meaning: visual and verbal semiotics in scientific text." In J. R. Martin & R. Veel (Eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science* (pp. 87–113). London: Routledge.

Mickens, Ronald E. (1990). *Mathematics and Science*. Singapore: World Scientific.

Newton, Isaac (1981). "Newton's simplified proof of fatio's reduction of the condition of the fall in least time along an arc of the brachistochrone to a curvature property of the cycloid [1700]." In D. T. Whiteside (Ed.), *The Mathematical Papers of Isaac Newton, Vol. VIII: 1697–1722* (pp. 86–91). Cambridge, UK: Cambridge University Press.

O'Halloran, Kay L. (1999a). "Interdependence, interaction and metaphor in multisemiotic texts." *Social Semiotics, 9* (3), 317–354.

O'Halloran, Kay L. (1999b). "Towards a systemic functional analysis of multisemiotic mathematics texts." *Semiotica, 124* (1/2), 1–29.

O'Halloran, Kay L. (2003a). "Intersemiosis in mathematics and science: grammatical metaphor and semiotic metaphor." In A.-M. Simon-Vandenbergen, L. Ravelli, & M. Taverniers (Eds.), *Grammatical Metaphor: Views from Systemic Functional Linguistics* (pp. 337–365). Amsterdam and Philadelphia: Benjamins.

O'Halloran, Kay L. (2003b). "Educational implications of mathematics as a multisemiotic discourse." In M. Anderson, A. Sáenz-Ludlow, S. Zellweger, & V. V. Cifarelli (Eds.), *Educational Perspectives on Mathematics as Semiosis: From Thinking to Interpreting to Knowing* (pp. 185–214). Ottawa: Legas Publishing.

O'Halloran, Kay L. (2004). *Multimodal Discourse Analysis: Systemic Functional Perspectives.* London: Continuum.

Oldershaw, Robert (1990). "Mathematics and natural philosophy." In R. E. Mickens (Ed.), *Mathematics and Science* (pp. 137–153). Singapore: World Scientific.

Reisch, Gregor (1496). *Margarita Philosophica*. Reproduced from 1535 reprint. John Rylands University Library of Manchester.

Rotman, Brian (1988). "Towards a semiotics of mathematics." *Semiotica, 72* (1/2), 1–35.

Rotman, Brian (2000). *Mathematics as Sign: Writing, Imagining, Counting*. Stanford, CA: Stanford University Press.

Shea, William R. (1991). *The Magic of Numbers and Motion: The Scientific Career of Rene Descartes.* Canton, MA: Watson Publishing International.

Stewart, James (1999). *Calculus* (4th ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Tartaglia, Nicolo (1546). *Quesiti, et Inventioni Diverse de Nicolo Tartalea Brisciano*. Venice.

Wigner, Eugene P. (1960). "The unreasonable effectiveness of mathematics in the natural sciences." *Communications on Pure and Applied Mathematics, 13*, 1–14. [Reprinted in Mickens 1990, 291–306.]

Chapter 6

Multimodality in language teaching CD-ROMs*

Martin Kaltenbacher
University of Salzburg, Austria

There is currently much discussion about the impact of new media and information technologies on learning and teaching. CD-ROM language 'trainers' are a relatively new type of resource which can potentially provide users with various new multimodal opportunities for language learning. This chapter demonstrates current strategies in such CDs for combining visualisations (sound waves, images and video-clips), sound (speech) and written text. While the purpose of these strategies is ostensibly to train the users' language skills and to help them understand the meaning choices in the target language, it will be shown that lack of relevant linguistic, semiotic or pedagogic expertise in the design process is too often delivering 'multimodality' which actually disrupts learning.

1. Introduction

In the introduction to his collection of essays on multimodality and multimediality, Baldry (2000: 22) states that "[all papers in the book] deal with multimodal texts as something more than a by-chance combination of written text and visual supports." Language teaching CD-ROMs integrate a variety of different media, thereby engaging a range of modes when training a user on particular lexical items, on grammatical structures, or on any of the four skills of reading, writing, speaking and listening. However, when critically investigating how different modes (e.g. text, image, sound, etc.) are combined in such CDs with the purpose to enhance the understanding and learning of certain linguistic meanings and skills, one often becomes aware that the combination of different modes is frequently not based on deliberate thought or on a didactic concept, but rather on chance.

This chapter discusses some of the most typical semiotic combinations that occur in language teaching CD-ROMs, and it critically evaluates whether and how the multimodal presentations enhance the understanding and learning of the meanings expressed through the different modes. It will be shown that different

semiotic encodings are often combined so ineffectively by the software designers that understanding or learning the intended meaning may not be possible.

2. Premises

The main advantage of a CD-ROM is that it combines various types of learning materials in one source. Text can be supported by sound and images, dialogues can be presented in the form of audio or video sequences, explanations can be followed up with interactive comprehension and production exercises, learners can record and listen to their own speech and compare it to a model speaker's utterances. These are just a few of the many possibilities for exploiting different modes in order to facilitate and speed up the user's learning process.

A decisive factor in the quality of such products is that the market for multimedia language learning materials is expanding with enormous speed. The amount of software made available by software providers is constantly growing. At the same time there is growing pressure on teachers at all types of educational institutions to make extensive use of IT and the new media in their classrooms. This is due both to the IT-industry's lobbying of governmental and educational authorities to spend more money on training pupils and students on computer-based products, as well as the wide-spread public belief that if something is 'multimedia', it is also automatically good.

Industry has responded immediately to the increased demand for such products. Software manufacturers and publishing companies have been quick in taking the opportunity to make good money in this large and fairly new market. Fierce competition has led to the negative effect that the dominant factor in the development of such software is speed rather than research and expertise. In addition, prospective customers are young and computer literate. Thus, most of the products target the so-called 'fun and play generation'. The new pseudo-pedagogical demand that all learning must be fun is all too visible in many of the products. More production effort seems to be put into designing the 'packaging' rather than the 'contents'. Games and sophisticated visualisations are so dominant in many of the CDs that the industry has coined the label 'edutainment' for the new materials.

A major challenge to good language teaching software is the fact that language is basically a creative and interactive tool for what, for example, Long (1996: 414) calls "negotiating for meaning". What lies behind this phrase is the concept that language used in whatever form (spoken, written, signed, etc.) between two or more interactants (speakers, readers, viewers, etc.) is modified until all persons involved know what each participant wants to convey. This functional concept of language use has been the fundamental principle in most language teaching

materials since the late 1970s (see Galloway 1993; Johnson 2001:182ff.). An immediate suspicion arises as to whether computer software can be a suitable tool for providing such creative and interactional language.

Up to now there is (to my knowledge) no software available on the market that can react to a learner's mis- or non-understanding by linguistically negotiating the intended meaning, and it seems unlikely that such software will ever be available. This is the point where non-verbal semiotic modes might help to clarify the meaning of an otherwise opaque linguistic sign. When a learner in a traditional classroom setting does not understand the word *dog*, for instance, s/he can ask the human teacher *What is a dog?*, and the teacher will usually come up with the proper explanation. A computer will remain silent on this question, no matter how vigorously you shout it at the screen. What a computer program can do instead is supply other semiotic encodings of the lexical item, either straight away or upon request. Possible modes for negotiating the requested meaning in such a case range from the very simple to the very sophisticated. Additional information could be presented in form of an extra window giving a simple translation of *dog*, (e.g. *Hund, chien, cane*, etc.). It could be visualised in the form of an icon or a photograph of a dog, which could again be accompanied by sound (e.g. barking). It could even be presented in video format. The range of possibilities to promote understanding is wide.

Such additional encodings must meet two basic requirements, a semiotic one and an economic one. The *semiotic* one requires that the additional encoding identifies the intended meaning (the signified, in the Saussurean sense) as clearly and exclusively as possible, e.g. by showing the picture of a dog and not just any animal. This might appear to go without saying, yet it is not necessarily obvious to software designers, as some examples below will show. The *economic* requirement is that the intended meaning should be conveyed to the user at the lowest possible technological cost. Naturally, a language teaching CD cannot provide a video sequence with every lexical item. Very often it cannot even offer an image with every item. Apart from technical restrictions to a CD's storage capacity this would not be communicatively functional either. To answer each comprehension check of a learner with a flood of new informational input would infringe Grice's (1989) cooperative principle.

Language teaching CD-ROMs come in three different categories. General English courses make up the biggest group of products available on the market. The second category contains CDs that concentrate on a particular skill or a particular aspect of language learning. Typical product names here are 'grammar trainer', 'vocabulary trainer', 'communication trainer', 'writing assistant'. The third category offers products that teach an LSP variety, most commonly Business English.

3. Examples of visual/audible 'learning aids'

Let us look in detail at some combinations of different modes used in typical products.

3.1 Pronunciation and acoustic phonetics

Many of the language teaching CDs available on the market contain some type of pronunciation exercise, where the learner can listen to single words or complete sentences produced by a model native speaker, record her/his own version and then compare the two versions. The producers of *Kommunikationstrainer Englisch 3.0* by Digital Publishing (1999) offer special training in the English vowels and consonants in what they call their "pronunciation lab". However, they seem to confuse graphemes with phonemes, when they train users to 'pronounce' letters of the English alphabet: "Besondere Beachtung bei der Aussprache finden die englischen Buchstaben g, h, j, v, w, y." (pronunciation of the English letters g, h, j, v, w, y needs special emphasis) (*Kommunikationstrainer Englisch 3.0:* Pronunciation, Vowels, Exercise 1). Incidentally, they also seem to confuse vowels with consonants.

It has been shown that the drilling of pronunciation patterns is pointless if the learner does not also receive explicit pronunciation instruction (e.g. Verhofstadt 2002: 121). Moreover, the drilling of isolated words or syllables, let alone letters, neglects the importance of prosody and intonation for natural speech.

A visual encoding from the field of acoustic phonetics that has become a widely used tool in language teaching CD-ROMs for checking the learners' accuracy in pronunciation is EVF – electronic visual feedback (Verhofstadt 2002: 128). EVF usually shows an oscillogram (a sound wave) of the drilled phrase as uttered by a model speaker and by the learner. Some programs add to this a numeric scale giving the percentage of correspondence between the two utterances. An example of this is given in Figure 1, taken from *Vokabeltrainer English* (Sprachlabor, Tricky words, Aussprache). The sound wave to the left represents the utterance of the word *except* produced by the American English native speaker serving as

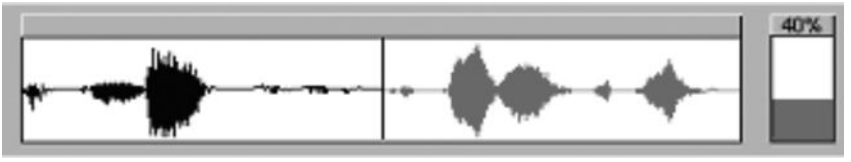


Figure 1. Sound waves for *except* generated by *Vokabeltrainer English*

the software's model speaker, the sound wave to the right is my production. The correspondence between the two versions is 40%.

When we consider the semiotic functions of this representation, it becomes clear that the choice of a sound wave as a tool for enhancing learning is a bad one. First of all, the average language learner has no training in acoustic phonetics nor in interpreting sound waves, and products making use of such representations do generally not provide such training either. Second, even if information on how to read a sound wave was provided, it would not be useful, as there is general agreement among phoneticians that a sound wave does not contain much phonetic, let alone any articulatory, information (e.g. Ladefoged 1996: 43). Third, displaying 'percentages of correspondence' between the two utterances is discouraging for the learner, as no information is provided as to where in one's particular articulation the lack of correspondence lies. The display offers no clue about whether the non-correspondence is due to inaccuracy in voicing, vowel length, stress, aspiration, place of articulation, etc. The only way for a learner to improve her/his pronunciation, or more precisely, to achieve a higher degree of correspondence with the model articulation, is through trial and error. In fact, even native speakers of English tested in this 'pronunciation lab' could not achieve results that were assessed as satisfactory by the software.

Summarising, we have to conclude that such EVFs are not suitable tools for teaching pronunciation to non-expert learners. The best they can do is to serve as eye-catchers: they are colourful, sophisticated and impressive. The only information a learner can get from them is that her/his utterance is not like that of the model speaker and therefore unacceptable, even if the latter is not the case at all.

3.2 Text and image

This section will look at a multimodal combination which is very frequent in language teaching CD-ROMs: the text + image combination. Images are used in language teaching CDs in two different ways. One is to show the typical situational context for a structure to be learnt. The user sees, for instance, the picture of a restaurant and is then presented with the verbal items to be learned, i.e. the different people and objects typically found in a restaurant, like *waiter, white wine, fish*. The other is to depict the very item to be learnt, establishing cohesion across several modes, i.e. a written word, a visual corresponding to its meaning, and often also a recording of the spoken word. In this case the "images have the function of illustrating an argument carried by a written word, that is of presenting ('translating') the contents of the written language in a different medium" (Kress & van Leeuwen 1996: 38).

Translated into the systemic functional terms laid out by Halliday (1994: 296ff.), each visualisation contributes to the information units in the different sections of a CD. Sections are organised through the arrangement of already known (Given) elements and some elements which are New and supposed to be learned. The first type of image represents the Given (see e.g. *restaurant* in Figure 2 below). It reinforces the context in which the new words and phrases (*waitress, white wine, dinner,* etc.) are presented. The second type of image depicts these New element (see e.g. *dinner* in Figure 5).

An important aspect is that in multimodal, just as in monomodal verbal text, "discourse has to start somewhere, so there can [must] be discourse-initiating units consisting of a New element only" (Halliday 1994: 296). This is achieved in language teaching CDs by arranging the material into different units entitled, e.g. *in the restaurant*, *at the hotel reception, at the railway station*. Entry ports to such units are usually highlighted textually and visually, and sometimes by means of short video clips and sound files.

So far so good – but a picture paints a thousand words. Depicting the isolated meaning of such a word as *dinner* poses some difficulties even within a context of *in a restaurant* or *eating.* One of the problems will be making explicit the distinction between denotative and connotative meanings of *dinner* in cultural context. What *dinner* means for me might be unacceptable (if not revolting) for somebody else. A further problem is that a complex picture, like Figure 5, may denote many separate meanings ('paint a thousand words'), which taken together may or may not raise in the learner a connotation that resembles the denotation of the lexical meaning of *dinner,* which the picture wants to convey. This problem of finding not one but many denotative meanings within a picture could easily be reduced by using icons instead of photographs or complex images, as icons are usually designed to denote one thing only.

All images in a CD fulfil a dual function. A photograph, drawing, cartoon or icon always serves as an eye-catcher to make textual information more attractive. The more important function, however, is to identify to the user the meaning of the verbal in a non-verbal semiotic code. As Wahlster (1996: 9.3) puts it, "an important synergetic use of multimodality in systems generating text-picture combinations is the disambiguation of referring expressions." Successful disambiguation depends on establishing appropriate cohesion between the verbal text and the picture. Wherever such cohesion is lacking, the text-picture combination is not useful for language teaching purposes, as the different modes convey non-identical meanings. Unfortunately, CD-ROMs frequently contain pictures that do not disambiguate the meanings of the verbal text. In many cases the pictures are either too general or they are ambiguous, so that they allow more than one semiotic

interpretation. Even worse, they sometimes encode a meaning that is completely different from that in the text.

The following examples shall illustrate this claim. They are taken from the CD *Grosser CD-ROM Sprachtrainer Englisch* (published by Tandem Verlag, Multimedia Line 1995) but are symptomatic of many digital language teaching products available on the market. This CD-ROM is designed as a holiday vocabulary trainer for beginning learners of English as a foreign language and systematically explores a range of fields that a tourist may need to be familiar with when travelling abroad, e.g. *hotel, restaurant, pharmacy, hospital, post office,* etc. The different sections provide learners with lists of relevant lexical items as well as lists of sentences that may be useful in each of these contexts. All the meanings are encoded in three different modes: first, written English text plus German translation; second, a photograph accompanying each isolated word and a situational cartoon accompanying each sentence; and third, a recording of the English word or sentence spoken by a native speaker. In addition, the learner can record him/herself and listen to this recording and the model speaker's version as often as desired. The dominant mode in each presentation is the image, which is centrally arranged on the screen under the rather small verbal text of the word/sentence given in German and in English. These images only rarely encode the same meaning as the written text. The following examples are taken from the field *restaurant*.

We enter the section with a mouse-click on the picture of a restaurant in the main menu and find ourselves inside an empty restaurant, see Figure 2. This photograph presents a New discourse-initiating point that establishes the context

Figure 2. Empty restaurant (full of 'spooky' voices)

(Given) for all the words and sentences that one can access within the section. A surprise here is that the accompanying sound does not match the meaning of the visual. The photograph shows an empty dining room with tables nicely set with posh china and silver, ready to accept a flood of guests. The sound, however, is that of a full restaurant with people talking and laughing, cutlery clattering and glasses touching one another, as toasts are proposed.

A mouse-click on the menu lying on the front table opens a sub-section of New elements contained in a long list of different types of meals, dishes, and food, headed by the three words *breakfast, lunch* and *dinner.* Another click on these words generates the pictures in Figures 3, 4 and 5.

Figure 3. Frühstück/breakfast

Figure 4. Mittagessen/lunch

Figure 5. Abendessen/dinner

None of the three photographs can be called a good choice for the lexical meanings they have been chosen to identify. Apart from the poor picture quality the image in Figure 3 is too general. We see a typical restaurant or cafeteria situation with one waiter and two guests, but there are no contextual clues that the couple is having *breakfast* rather than *tea, coffee, lunch, snack, supper, ice-cream* or anything else. The food that is consumed by the guests is not identifiable, and the picture lacks a time reference to 'morning', which could easily be given in form of a clock or people wearing pyjamas. While the image in Figure 3 is too general, the picture in Figure 4 (a cup of coffee and three half croissants) is simply inappropriate to both the German as well as the Anglo-American concept of *lunch*. An important issue is that food and eating habits vary considerably across different cultures and languages, and so there is a nearly endless variety of possible connotations for individual learners. One connotation of *Mittagessen* for Austrians and Germans, for instance, is that it is the main meal of the day.[1] In Britain, however, lunch would typically be a snack consisting of sandwiches, and dinner would be the main meal of the day. In neither culture does lunch ever consist of coffee and croissants.[2] For the same reason Figure 5 is not well chosen to identify the cross-culturally different meanings of *dinner.* The image depicts a three-course meal consisting of soup, a main course of meat served with a side dish and dessert. Again, it is worthwhile noting that the picture does not denote one thing but many, the sum of which may connote 'main meal of day', which makes it ambiguous, since this would typically be *dinner* in the U.K. but *lunch* in many Austrian and German contexts. In all three pictures in Figures 3–5 a time indicator, such as a clock, would have been an easy solution to the problem.

Figure 6. Couple in restaurant, man with female voice (eunuch, counter tenor, transsexual?) shouting for a *waiter/waitress*

Grosser CD-ROM Sprachtrainer also trains learners on a number of sentences with so-called complementary lexical meanings (Kastovsky 1982; Lyons 1995), e.g. *dead* versus *alive, married* versus *unmarried*. The essence of teaching such pairs is to distinguish the opposite meanings from one another. It is therefore particularly surprising to find CD-ROMs using the same visual to encode both. The picture in Figure 6 shows a couple in a restaurant, the man is apparently shouting. The accompanying verbal text is *Herr Ober / waiter*. At best, this picture signifies contextual information, serving as a reiteration of the Given, as it repeats the information that this scene is taking place in a restaurant. However, the same cartoon also accompanies the complementary lexeme *Fräulein/waitress*. The only difference is that the verbal text in the title of the picture now encodes a complementary meaning. What is missing is the essential lexical-visual cohesion between the verbal and the visual elements. This could easily be established by portraying the New elements of the complementary lexical items, i.e. one male and one female restaurant employee, in two separate cartoons. What is also odd, though rather amusing, is that the spoken texts *waiter* and *waitress* have been recorded with a female voice, though the picture shows them as being shouted by a man. In fact, all the sound files on this CD are recorded by the same female speaker, although they accompany images which variously represent female and male persons. This is unacceptable in terms of grammatical cohesion across the different modes, if we link gender to grammar here. Equivalent verbal nonsense would be a sentence like **He bought herself a new car.*

Figure 7. Request for a table in a smoking area

Figure 8. Request for a table in a no-smoking area

Similar problems occur in the pictures supporting the two complementary sentences *Could we have a table in the smoking/no-smoking area?*, presented here as Figures 7 and 8. Again there is no element in either picture that coheres with *smoking* or *no-smoking*. The complementary meanings encoded in the two images are: *a man talking to a waiter* (Figure 7) and *a woman talking to a waiter* (Figure 8). The intended meanings could have been made clear by showing a restaurant

Figure 9. Lady inquiring about *something to eat, low fat dishes, soup, fish, white wine*

with two separate areas, one with smokers, one without, and the guests pointing to the respective area in which they want to sit or by showing a smoking couple in one image. Both pictures again reflect only the Given context and neglect the New information, which is the essential one for the learner.

Figure 9 shows the same couple sitting at a table, the woman talking to the attentive waiter. Out of the 20 sentences that are presented as 'useful phrases when going to a restaurant', this image accompanies five. Each of these sentences has a different meaning: *I'd like something to eat. Do you have any low fat dishes? I'd like some soup. I'd like fish. I'd like white wine.* Listing all the New information in the clause final positions, we have: *something to eat, low fat dishes, soup, fish, white wine.* Similarly to Figures 7 and 8, none of this New information is encoded in the image. The visual meaning is too general, and the picture is therefore superfluous. It can only serve as an eye-catcher to attract the learner's attention but does not help her/him understand what these sentences mean.

In contrast, let us have a brief glance at Figure 10. It presents the couple at the end of the meal. The man is talking to the waiter and pulling a plastic card out of his suit, showing it to the waiter. From this it is easy to infer that the man is enquiring about credit cards. The sentence supported by the picture is: *Do you take credit cards?* This is one of the good multimodal combinations in the CD, where the two meanings encoded in the verbal (spoken and written) modes and the visual mode match, i.e. there is semiotic cohesion between the phrase *credit card* and the yellow plastic card shown in the man's hands. The picture at the same time reiterates the Given context (scene in a restaurant) and portrays the New

Figure 10. Man inquiring about *taking credit cards*

information (credit card) in a central, foregrounded position. The only remaining problem is that the man still speaks with a female voice.

Summarising, the visual semiotics in many of the images found in language teaching CDs are not helpful to the learners in perceiving the flow of information from Given to New. In other words, the images do not reinforce the verbal meanings to be learnt, but those which are already known.

3.3 Speech and gesture / lip movement

The third type of visualisation that will be addressed here is the short video-clip used for vocabulary and pronunciation training. It usually lasts between half a second and two seconds and shows native speakers producing words or short phrases in the target language. Such video-clips frequently present close-ups of the model speaker's upper body and/or face, and focus on gestures, hand-movement, head-movement, eye-contact, etc. The most important feature is, however, the lip movements.

The following are examples of this kind of visual used in the language learning CD-ROM *Sag's auf Englisch* (*Say it in English*) by Langenscheidt Publishing Company (1998). This CD, too, is promoted as a holiday trainer, teaching users the basic words and phrases for tourists. Each word or phrase on the CD is accompanied by a very well-chosen icon supporting its meaning. These images are kept at a very simple level, and due to their iconicity they mostly avoid the lexical ambiguities and wide range of possible connotations that rest within more complex

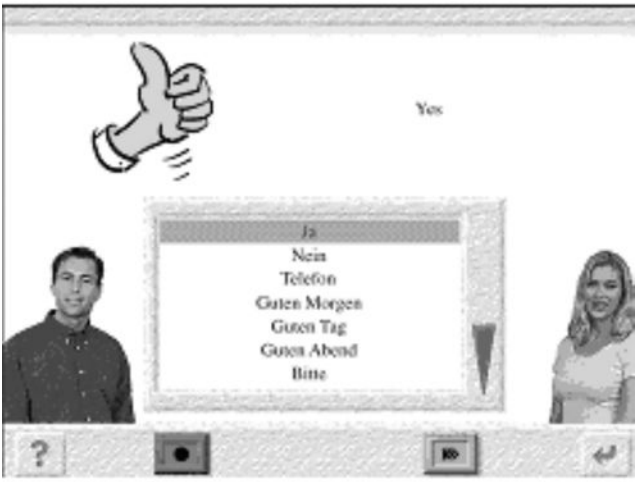


Figure 11. Frame from the video-clip showing speakers at the end of uttering *yes*

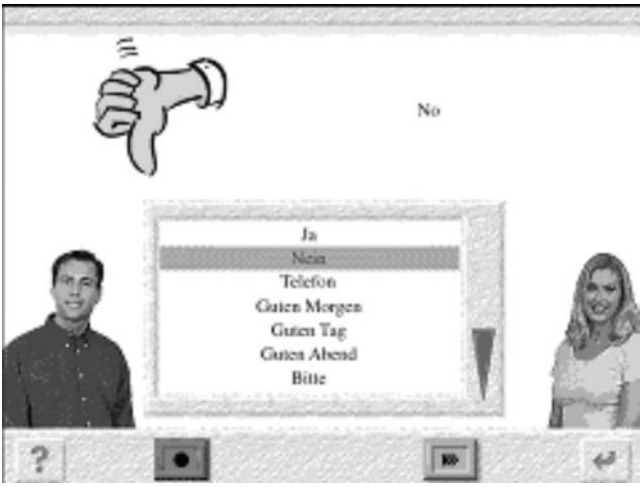


Figure 12. Frame from the video-clip showing speakers at the end of uttering *no*

pictures. Two examples are the two hands with the thumbs pointing up to signify *yes* and down to mean *no* in Figures 11 and 12. In addition to that, each word or phrase is also accompanied by two short video-clips alternately showing a man or a woman uttering the word/phrase. And this is where the problems start. Figures 11 and 12 each display one frame from the clips showing the two model speakers uttering the complementary words *yes* (11) and *no* (12) respectively. Scrolling through the wordlists to listen to different words and phrases, it is surprising to

find that all the model productions are supported by the same two video-clips. In other words, CD contains only two video sequences, one for all the man's utterances and one for all the woman's. This is of course a ridiculous strategy – as Bellik (1996: 9.4) points out, "in human communication, the use of speech and gestures is completely coordinated". Gesture is a mode with its own functions in spoken interaction – raising an eye-brow may express doubt or curiosity, particular hand movements lay emphasis on certain aspects of an utterance, etc. If only one gesture accompanies all utterances (some a single word, others extended phrases or sentences), it will not help learners understand diverse meanings. In these video-clips, the speakers are shown in a relatively static posture – standing or sitting upright without hand movement, both gazing out of the screen, establishing eye-contact with the learner. While the man nods his head, the woman waggles her shoulders slightly. These are the only gestures in the videos. The learner is therefore confronted with the apparent message that all English words, phrases, and sentences are accompanied by exactly the same gestures. If you are female you wriggle your shoulders, if you are male you nod your head, whether the verbal message is *yes, no, telephone, sorry, lady's toilet, gent's toilet, left, right, beer, please speak more slowly, where can I hire a car, is it safe to swim here, cheers* or any other of the hundreds of utterances accompanied by these two gestures.

But the worst effect of this strategy is that with each speech segment the learner sees a physiologically incorrect articulation. In Figures 11 and 12 identical lip movements are seen for *yes* and *no*. The position of the male speaker's lips is impossible for producing /nəʊ/, which is characterised by lip rounding. It is common knowledge among linguists that in spoken interaction "the optical information is complementary to the acoustic information because many of the phones that are said to be close to each other acoustically are very distant from each other visually" (Goldschen 1996: 9.5). Language learners can be expected to focus on the lip movements of teachers for disambiguation and as a guide for their own articulations. Consequently, this kind of digital 'teacher' is likely to be less help than no visual guide at all. Benoit et al. (1996: 9.6) point out that "[...] synthetic faces increase the intelligibility of synthetic speech, but under the condition that facial gestures and speech sounds are coherent." This also holds true for visual recordings of real human speakers. It is a fact that people may perceive the same syllable or word differently if accompanied by misleading visual data, which is generally known as the 'McGurk effect'. In a groundbreaking experiment McGurk and MacDonald (1976) showed a manipulated video to their test persons, in which they heard the syllable /bɑ/ while seeing a speaker utter the syllable /gɑ/. As a result of this multimodal incohesion, their subjects all reported that they were perceiving /dɑ/. We must therefore conclude that a video-clip with visual articulations that

do not correspond to the recorded utterances is likely to result in poor learning of the target language.

4. Conclusion

Over the last ten years there has been a growing demand by linguists as well as other experts on semiotics to acknowledge equal status to the visual as to the verbal semiotics in multimodal texts. This demand has been amply documented, e.g. in Kress and van Leeuwen (1990, 1996, 2001), Baldry (2000), O'Halloran (2004) and in this volume. One can easily foresee that theoretical as well as applied issues of multimodal semiotics will continue to move into the focus of linguistics, semiotics, communication studies and other sciences over the next years, and it is high time that this be so.

The main advantage of new information technology and media is the relative ease of use and the speed with which new multimodal texts can be designed and distributed. Ease and speed, however, both increase the risk that authors fail to do their research or to consult existing expertise in the design process for multimodal products. Unfortunately, this seems to be so in many cases in the language teaching CD-ROM industry. The majority of such materials are not based on a solid foundation of semiotics, linguistics, language acquisition theory, or pedagogy. In fact, they revive many approaches common in the 1950s and 60s, such as behaviourist pattern drilling, which have otherwise long been discredited and dropped from language teaching practice (for a fuller discussion, see e.g. Kaltenbacher 2003; Ventola & Kaltenbacher 2003; Chapelle 1997; Holland et al. 1995). The underlying principle in many language learning CDs seems to be "We now have the technology, so let's have a go!" One evident design criterion is economics – the software should be quick, easy, and cheap to programme. Another is entertainment – visual media (e.g. sound waves, video-clips) often serve as 'edutainment' gimmicks without providing real help to learners. Producers mainly focus on technical software expertise and ignore the fact that there is more to producing a complex multimodal text for teaching a language than simply putting words and images together.

Notes

* I would like to thank Digital Publishing and Tandem Verlag for their explicit permission to reproduce picture material from the CDs *Vokabeltrainer 3* and *Holiday Trainer English*. Equally, I would like to thank Langenscheidt KG for their permission to reproduce material under

the provisions of the Fair Use Paragraphs of international copyright law (e.g. Österreichisches Urheberrechtsgesetz §§40, 46, US copyright law 17 USC §107).

1. This has been changing over the last years, especially in urban areas, where many people have a snack for lunch and the main meal in the evening. Families with children, however, and most people in the country still have the main meal of the day at noon and a light supper in the evening.

2. The CD-ROM, designed for the German speaking market, was programmed in Kiev/Ukraine. However, a Ukrainian informant assured me that coffee and croissants would never pass for lunch in the Ukraine either.

References

Baldry, Anthony (2000). *Multimodality and Multimediality in the Distant Learning Age.* Campopasso: Palladino Editore.

Bellik, Yacine (1996). "Modality integration: speech and gesture." In R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenenand, & V. Zue (Eds.), *Survey of the State of the Art in Human Language Technology*, Ch. 9.4. Portland, OR: Center for Spoken Language Understanding. http://cslu.cse.ogi.edu/ HLTsurvey/HLTsurvey.html

Benoit, Christian, Massaro, Dominic W., & Cohen, Michael M. (1996). "Modality integration: Facial movement & speech synthesis." In R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenenand, & V. Zue (Eds.), *Survey of the State of the Art in Human Language Technology,* Ch. 9.6. Portland, OR: Center for Spoken Language Understanding. http://cslu.cse.ogi.edu/HLTsurvey/ HLTsurvey.html

Chapelle, Carol A. (1997). "Call in the year 2000: Still in search of research paradigms?" *Language Learning & Technology, 1* (1), 19–43.

Galloway, Ann (1993). "Communicative language teaching: an introduction and sample activities." Washington, DC: ERIC Clearinghouse on Languages and Linguistics. http://www.ed.gov/databases/ERIC_Digests/ed357642.html

Goldschen, Alan J. (1996). "Modality integration: facial movement & speech recognition." In R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenenand, & V. Zue (Eds.), *Survey of the State of the Art in Human Language Technology,* Ch. 9.5. Portland, OR: Center for Spoken Language Understanding. http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey. html

Grice, Paul (1989). *Studies in the Way of Words.* Harvard: Harvard University Press.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar.* London: Edward Arnold.

Holland, Melissa V., Kaplan, Jonathan D., & Sams, Michelle R. (1995). *Intelligent Language Tutors: Theory Shaping Technology.* Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson, Keith (2001). *An Introduction to Foreign Language Learning and Teaching.* Harlow et al.: Longman.

Kaltenbacher, Martin (2003). "Language learning via CD-Rom – old wine in new bottles." In D. Newby (Ed.), *Mediating Between Theory and Practice in the Context of Different Learning Cultures and Languages* (pp. 171–175). Strasbourg: Council of Europe Publishing.

Kastovsky, Dieter (1982). *Wortbildung und Semantik.* Düsseldorf: Schwann-Bagel; Bern, München: Francke.

Kress, Gunther & van Leeuwen, Theo (1990). *Reading Images.* Victoria: Deakin University Press.

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images. The Grammar of Visual Design.* London and New York: Routledge.

Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication.* London: Edward Arnold.

Ladefoged, Peter (1996). *Elements of Acoustic Phonetics.* Chicago: University of Chicago Press.

Long, Michael. H. (1996). "The role of the linguistic environment in second language acquisition." In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413–468). San Diego: Academic Press.

Lyons, John (1995). *Linguistic Semantics. An Introduction.* Cambridge: Cambridge University Press.

McGurk, Harry & MacDonald, John (1976). "Hearing lips and seeing voices." *Nature, 264*, 746–748.

O'Halloran, Kay L. (2004). *Multimodal Discourse Analysis: Systemic Functional Perspectives.* London: Continuum.

Ventola, Eija & Kaltenbacher, Martin (2003). "Lexicogrammar and language teaching materials – a social semiotic and discourse perspective." In J. E. Joyce (Ed.), *Grammar in the Language Classroom* (pp. 158–201). Singapore: SEAMEO Regional Language Centre.

Verhofstadt, Katrien (2002). A Critical Analysis of Computer-Assisted Pronunciation Materials. University of Ghent, unpublished dissertation.

Wahlster, Wolfgang (1996). "Text and images." In R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenenand, & V. Zue (Eds.), *Survey of the State of the Art in Human Language Technology,* Ch. 9.3. Portland, OR: Center for Spoken Language Understanding. http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html

Reviewed language teaching CD-ROMS

Desktop Systems Ltd., Kiev, Ukraine (1995). *Holiday Language Trainer – Englisch; Grosser CD-ROM Sprachtrainer Englisch.* Königswinter: Tandem Verlag Multimedia Line.

Digital Publishing (1999). *English: Kommunikationstrainer 3+.* München: Digital Publishing.

Digital Publishing (1999). *English: Vokabeltrainer 3.* München: Digital Publishing.

Langenscheidt KG and EuroTalk Ltd. (1998). *Sag's auf Englisch.* Berlin and München: Langenscheidt.

Chapter 7

The multiple modes of *Dirty Dancing*

A cultural studies approach to multimodal discourse analysis

Markus Rheindorf
University of Vienna, Austria

The chapter provides a brief overview of my work on multimodality in film, while also focusing on multimodal signification in *Dirty Dancing* (1987) from the perspective of cultural studies. As such, it highlights the ways in which various modes contribute to the construction of specific meanings in the film's diegetic world. The chapter takes an instrumental view of multimodal discourse analysis, regarding it as a means rather than as an end in itself, and puts it to service for such issues as filmic genre and the performativity of class and gender.

The approach adopted here is furthermore transdisciplinary in the sense that it draws on various theoretical positions, concepts, and methods associated with the disciplines of linguistics, film studies, and cultural studies – at different levels of analysis. On the one hand, cultural studies provide a theoretical framework for the engagement with film as a cultural practice, semiotic and linguistic concepts and methods serve as the 'work horse', doing most of the actual analytical work. On the other hand, film studies are drawn upon, mostly for its longstanding experience with the medium, its practical insights into the processes of production and reception, as well as its previous cross-fertilisations with linguistics – mainly in the shape of Christian Metz's structuralist analysis of filmic codes.

1. Introduction

This chapter focuses primarily on the patterns of multimodal signification in the film *Dirty Dancing*, but is also concerned with describing the ensemble of modes specific to the medium film and in doing so suggests how a theory of multimodality can deepen our insights into the medium as well as its genres. The choice of this somewhat odd object of study in connection with multimodality has at least three separate, but ultimately interrelated, reasons. First, *Dirty Dancing* is *film*, a medium so obviously multimodal that film studies have long concerned them-

selves with the interplay between the visual and the auditory – although, of course, this rudimentary distinction of sensory 'modes' needs to be replaced with more subtle distinctions. Second, critics and fans alike have usually and conveniently labelled *Dirty Dancing* a 'dance film', and as such it does indeed depend on the use of dance as a social practice and a form of signification. Third, it is a film that is implicated in a host of cultural forces and thus bears witness to them. Both in the 1960's that the film purports to portray, as well as in the 1980's that produced it, the issue of class in American society was perhaps the most important of these. Thus, despite or maybe even because of its cliché-ridden story, *Dirty Dancing* is a surprisingly dense text recontextualizing complex social issues such as class and gender in a remarkable way: its cinematography employs characteristic distributions of modes in order to produce particular kinds of meaning.

Having said all that, it is perhaps necessary to point out the institutional context of the work this chapter grew out of and draws upon. It is primarily the representations of class in contemporary American culture – and of the social practices oriented around them – that are the interest of an on-going project conducted by a fluctuating group of graduate students and staff at the department of American Studies at the University of Vienna. Although this work for the most part involves a textual analysis of some kind, it should be pointed out that this is being done within the framework of a cultural studies approach.[1] This, then, was the immediate point of departure for engaging with *Dirty Dancing*, but it soon became evident that an analysis of the distribution of modes in this film (or, indeed, any other film) could well become crucial to the task. In order to test certain assumptions, other instances of dance films, such as *Flashdance* (1983), *Saturday Night Fever* (1977), *West Side Story* (1961), and *Grease* (1978), were later also included in the project. This broadening of the scope of the study, at least in part, also served to accommodate a growing interest in the notion of *genre*, various conceptualisations of which have haunted film studies for nearly a century. This problem remains largely unresolved to date. The idea was – and still is – to substantiate the vague concept of genre in film studies – the 'dance film' is just one example – with the help of multimodal discourse analysis.

In the course of the analyses, the following questions concerning multimodality have been paramount: (1) Is the distribution of semiotic modes in *Dirty Dancing* significant for the realisation of a generic structure – in this case, the 'dance film'?, (2) If so, how do various modes combine to realise this structure?, and (3) Can the 'dance film' be topologically related to other genres? In order to provide an interpretative framework for answering these questions, a number of methods have been tentatively combined to cover both film theory and semiotics in the wider sense. These include, among others, Metz's (1974) structuralist distinction of codes, Kress and van Leeuwen's (1996) framework for an analysis of visual el-

ements, van Leeuwen's (1999) account of sound and music, and Kress and van Leuuwen's (2001) more recent – but also more cautious – work on multimodality.

As these authors have repeatedly emphasised, meaning is always social.[2] It would seem that the meaning of dance in film is therefore equally dependent on both the cultural context and the background knowledge of the spectator. Despite this call for careful contextualism, it may, however, ultimately be possible to contribute meaningfully to the concept of genre as such by studying the workings and distributions of modes in genre films. With these issues on a larger scale on the theoretical horizon, this chapter focuses on a discussion of the findings with regard to *Dirty Dancing* and refers to issues of genre only where the analysis is likely to have more general implications.

2. Constructions of class

As far as the multimodal constructions of class from the perspective of cultural studies is concerned, *Dirty Dancing* can be regarded as an instance of what both filmmakers and those engaged in film studies have long realised: that *film* as a conventionalised mode of representation tends to reduce almost any sort of social conflict – and in particular conflicts of class, race, and gender – from a global to an individual level, whether that means an epic battle between good and evil or the kind of culture wars within a hegemonic society that are interesting from the point of view of cultural studies.[3] Thus, film can stage but also resolve class conflicts on an individual level, leaving the underlying social structure of its fictional world not only intact, but effectively unchallenged. And *Dirty Dancing* is no exception to the hegemonic discourse of 'Hollywood' or 'mainstream cinema', doing no more than allowing a handful of characters to contest class as a "structure of feeling", to use Raymond Williams' (1961) phrase, on a strictly localised level. Even in this narrowly defined space, however, most films will mask this conflict with one of personality, or find other suitable metaphors for the struggle with and against class boundaries. *Dance*, it would appear, is one of these alternative modes commonly employed by films – and it is one that allows for an apparent dissolution of boundaries on the dance-floor, a space that (like the sports-ground) is often conceived of in popular imagination as transgressing all class boundaries.

However, by highlighting professional opportunities for lower class people, this representational practice draws on – but also feeds into – the image of the lower classes as being somehow more 'physical', more 'body' and less 'mind' than the middle or upper classes. Thus, it is possible to make class visible and tangible through individual dancing bodies that move within a more or less closed system of dance as an indicator of social class. In many ways, *Dirty Dancing* depends on

this use of dance as a resource for the making of social meaning, as well as on the stigmatisation of the dancing bodies of the lower classes as being somehow uncivilised and *dirty* – both in the literal and a figurative sense.

3. Constructions of gender

Regarding the representation of *gender* in *Dirty Dancing*, it is helpful to follow Judith Butler in seeing gender as "always a doing, though not a doing by a subject who might be said to preexist the deed [. . .] identity is performatively constituted by the very 'expressions' that are said to be its results" (Butler 1990: 33). From this point of view, one can see the film as constructing gender as 'classed' – or, rather, as depicting 'gender performance' or 'performativity' as opposed to a stable and fixed category. It follows that other socially constituted and constitutive realities can potentially influence, subvert, and effectively reconstruct positions usually associated with gender performance. Among these, one may, of course, count aspects of *ethnicity* and, as in the case of *Dirty Dancing*, *class*.

In furthermore seeing the female body as "used exclusively as the metaphor for the corporeal side of the mind/body dualism", Gowaty (1997: 97) elaborates Butler's position by adding that the female body is seen by society as "ruled by emotions, standing in contrast to the masterful, masculine will, the locus of social power, rationality, and self-control". Significantly, however, the mind/body dualism referred to here is also at work in a classed society, distinguishing the lower classes – the figurative workforce and body of society – from its mind or head, i.e. the elite of a given society. While Gowaty's account of the dualism at work in gender role construction holds true as such, other social factors can be superimposed over a gendered power structure. Thus, in the relationship between Baby and Johnny in *Dirty Dancing*, gender is at times backgrounded and superseded by constructions of class which are in stark contrast to the conventional distribution of power along the male/female divide. Instead, at the crucial moments of the plot, it is the female but upper class body of Baby Houseman which is the site of will, self-control, and social power, while the male body – and one certainly gets to see a lot of Swayze's masculine upper body – is more or less at the mercy of the female.

This reversal of the relations of (social) power – i.e. the roles of defining and being defined as 'other' in the relational performance of gender – can also be discerned in Johnny's relationship to women in general, several of whom have paid him money for late night 'dance lessons'. Significantly, the film makes much of the fact that this is something that Johnny finds himself quite powerless to prevent or decline, something that frustrates rather than affirms his masculinity. After all, he is challenged by women who cannot only pay him for sex, but who could just as

easily have him fired – as one of them very nearly does – should he defy them or refuse their advancements.

Having said all that, it needs to be added that Johnny is only stripped of his masculinity in terms of gender performance when he finds himself outside the social context of his kind of people, that is to say the entertainment staff at Kellerman's. With them, he is something of a leader, obviously the best dancer, and commands considerable respect and social esteem. This is visualised or embodied most effectively in the many dance scenes of the film, which – whether one considers dance a spatial practice in the terms of de Certeau's (1985, 1988) semiotics of the everyday or as a mode of expression in the terms of Kress and van Leeuwen's (1996, 2001) theory of multimodality – show very clearly that he is given (or rather, is enabled by his status to take) the space that belongs to him.

Through dance, more than any other mode, *Dirty Dancing* establishes that Johnny's status is contingent on the social context in which he moves, acts, or dances. By using dance as mode of signification, the makers of the film were thus able to translate this metaphor smoothly into spatial terms, visually illustrating how Johnny and his partner Penny successfully use their dancing bodies to stake out their own territory, briefly, even among members of the upper class. Equally potent, however, are the images of their demotion when the true patriarch of Kellerman's – with a single flick of the hand – denies them the space they have filled with their dancing bodies, and from which they have drawn all their power and charm.

4. Uses of multimodal discourse analysis

Following an initial reading of the film intended to reveal the film's constructions of class as well as gender, a much more specific analysis of the multiple modes of the film was conducted, approaching speech, gesture, camera movement, image composition, soundtrack and dancing as semiotic resources.[4] The first questions, then, with regard to multimodality, were: (1) Which modes are employed in the film? (2) What is the relation between these distinct modes? (3) How do they converge to produce specific meanings?

As mentioned before, the approach taken here draws, among other methods of visual analysis, on the one developed by Kress and van Leeuwen (1996). However, any analysis conducted strictly within the confines of this method faces considerable difficulties when confronted with the sequential, time-based texts of moving pictures. On the other hand, methods developed specifically for the transcription of film – such as Thibault's (2000) frame-based multimodal transcription – tend to be unwieldy and all but impossible to implement with an entire feature film.

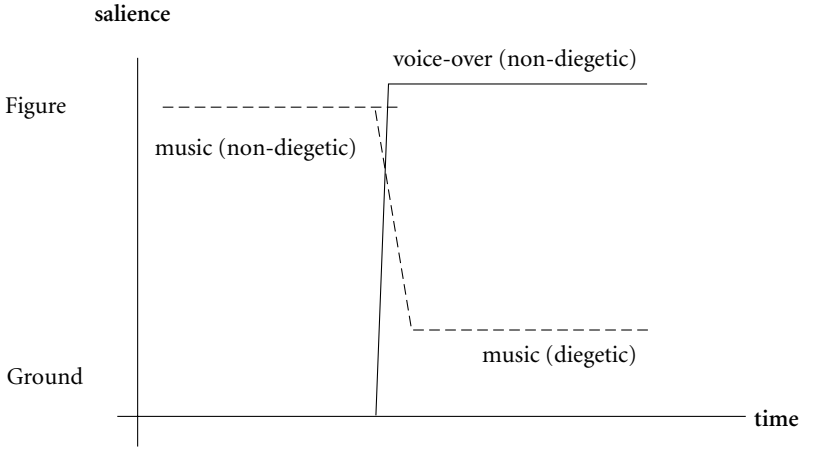


Figure 1. Example of graph showing modal changes

What is needed instead are more systematic ways of accounting for the deployment and unfolding of semiotic resources in the filmic medium. The following therefore draws on an array of other methods from the field of both semiotics and film studies – such as peak-and-trough graphs for the purpose of highlighting the changing deployment of specific resources or combinations of resources – in an attempt to construct such an alternative approach.

It is, however, instrumental to be also able to depict schematically the combinations themselves in order to relate them to the notion of media specificity as well as the concepts of diegetic and non-diegetic space. The first step, as it were, is a rudimentary and idealised map of the semiotic resources of film from the perspective of a multimodal discourse analysis, given in Figure 1. This map distinguishes between (1) the two major sensory modes of representation in film, that is visual and auditory semiotic resources, and (2) between resources that are either specific to film and have been conventionalised in and through their use and those which are non-specific to the medium and whose use has not been conventionalised in the history of film. In the context of this chapter, 'medium' is understood primarily as the material and technological properties of a given form of communication, as well as its technical possibilities and restrictions of production, distribution, and reception (Kress & van Leeuwen 2001: 22) – and thus distinct from the modes that have become associated with it and which Kress and van Leeuwen (2001: 22) see as conventionalised uses of a medium.

Certain difficulties, however, arise in relation with Kress and van Leeuwen's (2001) wording in their assertion that media can "become" modes through repetitive and conventionalised use. This would seem to imply that the medium through this 'abstraction' ceases to exist, or even that it cannot exist separately. Far from

denying the interdependence of media and modes, this chapter assumes that modes are indeed strongly associated with the media that have given rise to them. All the same, modes can be – and are – realised in more than one specific medium.[5] In addition, one can assume the existence of something like a semantic of a given medium, the field of all its possible uses, many of which may never be actualised, but are nevertheless possible. It is therefore neither practicable nor tenable to equate mode with medium, nor can they be seen as developmental stages of each other. What is at stake here is an opportunity to avoid the decade old trap of film studies' unhappy dependence on the notion of the 'essence' of film as a 'medium', while actually talking about its conventional uses. By separating the bare medium from its modes, one can speak of the properties of the medium while also saying something about its modes as historically and culturally determined uses.

Moreover, the distinction between specific and non-specific modes proposed here is often not a case of 'either – or'. In fact, they are often hard to separate and depend on other resources – systems of meaning, but not full-fledged modes – such as editing and camera movement. The boundaries around modes, in other words, show considerable permeability.[6] Furthermore, both specific and non-specific modes need not be stable and may change in the history of a medium, crossing boundaries, and may also vary from culture to culture.

The map given in Figure 1 is by no means meant to be an exhaustive description. Also, one should think of it specifically as a *map* as it does not provide an explanation to anything. Maps do not explain things. At their best, they can provide an overview, an idealised picture of things that may help one to get one's bearings. The grid formed by these initial distinctions can be seen in Figure 2, and located within its limits are the various modes associated with film.

The category of *linguistic resources* subsumes all those forms of expression that make use of language – and as uses of them can be found in all four areas formed by the two basic oppositions, they need to be placed so as to have a share of all four areas. It should also be noted that rather than use four isolated circles – for it is, of course, possible to consider spoken and written language separate modes – the one is intended to convey the relative continuity of the linguistic system as a resource shared by several modes. Furthermore, with regard to other semiotic modes, it was to be anticipated that the results of the analysis would sometimes indicate an overlapping or blurring of boundaries, and this can thus be represented as a move within a more general category rather than as a discontinuous jump between specifics. To illustrate briefly the uses of linguistic modes in film, there is: (1) the possibility to film non-specific written language, such as you would see, for example, on a warning sign, (2) the specific use of titles or subtitles, (3) the non-specific mode of spoken language as in dialogue (although certain forms of

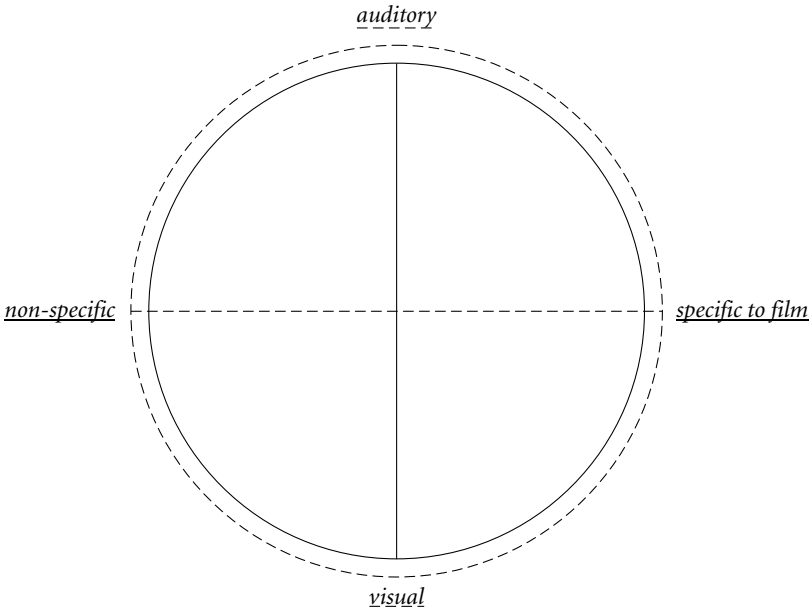


Figure 2. Basic distinctions of semiotic resources in film

dialogue may arguably be seen as specific) and (4) the film-specific use of spoken language, such as in narrative voice-over.

Drawing on van Leeuwen's (1999) work on auditory modes, the model proposed here also distinguishes between *speech*, *sound* and *music*. The latter category, too, can include both specific and non-specific uses. In the field of non-specific uses, for instance, one would find source music, while the most common film-specific example is the use of surround sound as mood music. Looking at actual instances of music in *Dirty Dancing*, however, one is faced with the problem of whether or not to include lyrics as a distinct mode, and if so, whether to place them as part of the linguistic or the musical mode. As there are strong arguments for and against either choice, the domain of lyrics is represented as an overlap between linguistic and musical modes.[7] A specific use of lyrics may occur, as it does in *Dirty Dancing*, when the lyrics of a particular song are used not only to support the action or plot (in which case they function on the same level as mood music), but to foreshadow or replace it. There is, of course, the distinct possibility that this may be possible only in certain genres such as the dance film or the musical.

Although sound obviously exists as a non-specific mode outside of the filmic medium, what is usually termed *sound effects* is available as a mode only to film. Distinct forms that have become established include, for instance, the whooshing sound that guns or other weapons in film apparently make when they are drawn

and the sound of space ships passing close by the point of view of the camera (even though sound cannot travel – and therefore in a sense does not exist – in outer space).

The focus of the analysis, as far as the visual modes were concerned, was on the meaningful uses of *dance*. These, it would appear, can be considered film-specific when choreographed for a particular camera perspective and/or movement, such as a top-down perspective, or when choreographies have developed for or through a particular genre, such as the musicals. It is important to note that the simple fact that dancing is of course edited to look 'better' than it would in real life (that is, more fluent and energetic than it would from one relatively fixed point of view), can also add to the meaning of the dancing, but would rely on the film-specific resources of *editing* and *camera movement*.[8] The dancing of dance films like *Dirty Dancing* is, at least in part, specific to film or maybe even to the genre. Consider, for instance, the dramatic movement in unison (and thus of solidarity and unity) within a group that is represented visually and in a conventionalised way in dance films (and many musicals) by a particular kind of dance choreography similar to the one at the end of *Dirty Dancing*. In addition, though this is not undertaken here, the relation of dancing to such modes as *movement*, *gaze*, and *gesture* will in the future need to be explored as well.

For a brief illustration of how the map proposed here can be used in conjunction with a multimodal analysis of individual scenes, consider the rather conventional opening scene of *Dirty Dancing* in its use of music and voice-over narration. As the car moves from left to right (from the given into the new, in terms of Kress & van Leeuwen 1996), it first opens up the diegetic space of the filmic narrative. At the same time the music, which initially began as non-diegetic 'Figure' and the focus of interest, is pushed back to the position of 'Ground' (that is, setting or context), as Baby's voice-over narration takes over as 'Figure'. The transitions are smooth and effective in their division of labour, masking the fractured space of the shots and creating a continuous space-time through the process known in film theory as *suture*. The changing distribution of specific and non-specific modes in this scene is characteristic of filmic openings, and all modal conventions have been adhered to.

As in the first sequence of the film, the makers of *Dirty Dancing* succeed at blurring the boundaries between specific and non-specific uses of music in several ways. For one, the distinction between source and surround soundtrack, or else between diegetic and non-diegetic sound, is constantly crossed. In fact, the film has been said to be almost as much 'about' music as it is about dancing.[9] Indeed, its makers are careful never to employ music simply as non-diegetic mood music, so that even when there is surround music, the film's mise-en-scéne always provides a record player or radio as the diegetic source and thus anchors or con-

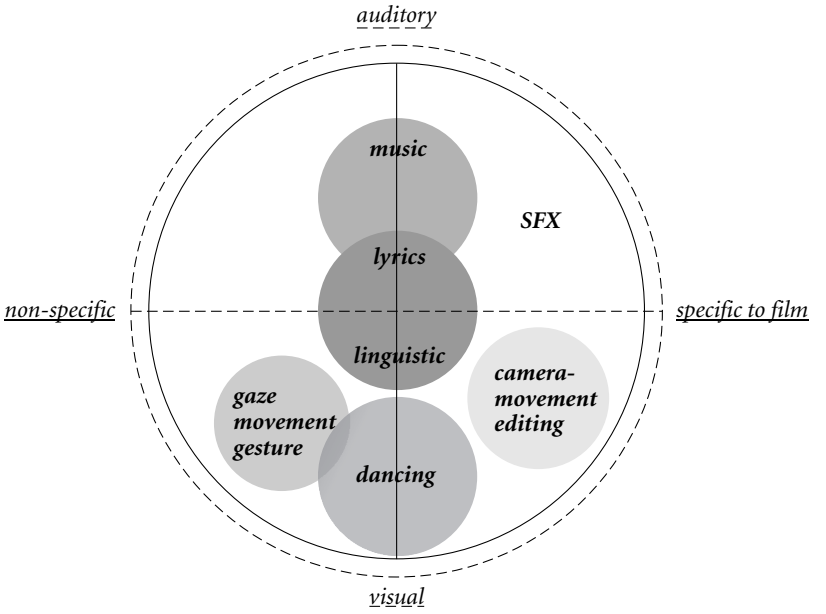


Figure 3. A map of semiotic modes in film

tains the soundtrack in the diegetic space of the film's narrative world. Repeatedly throughout the film, a particular track begins to play as diegetic sound (marked as such by its quality and distance) – and, in terms of van Leeuwen's (1999) approach, as 'Ground' – and then rises in volume and quality to become more than simply 'Figure' or the focus of interest, it in fact replaces all other sound. In other words, in becoming 'Figure' music in *Dirty Dancing* always also becomes 'surround' and thus non-diegetic. As van Leeuwen (1999: 170–181) put it in his work on the uses of soundtrack, it is one of the characteristics of sound that it is dynamic and can move the listener towards or away from a certain position. In the case of these tracks in *Dirty Dancing*, the music gradually opens up into and envelops the non-diegetic space of the world of the spectator. In one instance, as Baby and Johnny dance and sing along to the non-diegetic music track, the lyrics even come to function as a substitute for dialogue.

As has already been suggested, the boundaries between gesture and dance are prone to a different kind of blurring, so that in *Dirty Dancing* a single unbroken movement can frequently be seen as crossing the boundaries between the two modes. Like the gradual change of music from diegetic to non-diegetic, this has a similar effect in anchoring the beginning of sometimes choreographed and clearly stylised dancing in the diegetic space of the film (coinciding with a move from

non-specific to specific modes), while also opening up into the non-diegetic as the scene unfolds.

5. Multimodal configurations and issues of genre

In light of the distinct patterns that emerged during the analysis of individual scenes, it seems possible to extend Sinclair's (1991) assumption that language contains many prefabricated items such as set phrases and collocations to other modes besides language. Using multimodal discourse analysis, these 'filmic phrases' or sequences can be shown to share common patterns and distributions of modes,[10] among them the absence of linguistic information in certain contexts and in the case of certain content. Thus, it is possible to identify scenes with voice-over narration serving to launch the narrative of a film according to their characteristic configuration of modes as well as according to their content and function.

Moreover, film conventionally and strategically employs changing multimodal configurations in order to increase the salience of certain scenes. These techniques are central to a film's ability of endowing the ordinary incidents of everyday life with symbolic meanings. In fact, it is well-known in film studies that this is commonly achieved through the use of soundtrack, lighting, and camera movement – yet, a multimodal perspective can offer new insights into the phenomenon of salience as well as providing innovative ways of representing the choices made in such instances of meaning-making. A particularly common example of such a characteristic (and possibly generic) distribution of modes is what is usually known as the 'training or preparation sequence'. Such a sequence involves the depiction of a relatively long period of time in which the characters undergo some form of development, often involving physical training or other forms of preparation for a conflict that lies ahead.[11] In terms of multimodality, these scenes are conventionally realised through the use of (1) short cuts as far as the system of editing is concerned and (2) the use of non-diegetic music that is both 'Field' and 'surround'. The significant element in this case is not only the distribution of modes as a structure, but its change in the course of the sequence, that is the temporal unfolding of that structure or distribution. It seems that a combination of schematic maps and peak-and-trough graphs is the best method for visualising the structure and its unfolding respectively. Herein lies the crucial opportunity for multimodal discourse analysis to contribute to studies of genre, not only in film, but in every medium that employs more than one mode.

Apart from being remarkable for its cultural implications, the music in *Dirty Dancing* in some cases is even more remarkable for being typical music of the 1980's. This phenomenon coincides with the uses of music that involve the kind

of move from diegetic to non-diegetic sound that has already been discussed. The movie, though ostensibly set in the 1960's, was of course designed and produced in the 1980's – as the design and colour of the title readily testify. This historical and cultural distance between setting and spectator is bridged by a variety of means in *Dirty Dancing*, including Baby's narrative voice-over. On a different level, the design of the opening titles serves the same function, and so do the language and clothes of the eighties that pop up incongruently now and then in the film.

Even more prominently, however, the music of the 1980's sometimes intrudes upon the otherwise authentic 1960's soundtrack of the film.[12] Thus, the music tracks opening up into non-diegetic space were in nearly all cases also recognisably made in the 1980's, even if the remainder of the songs featured in the film perfectly fits the year 1963.[13] Nevertheless, even in these cases of incongruence, the film locates the source of the music as it begins to play in the diegetic space of the film, and only then allows it to open up into the non-diegetic. It would appear, then, that it is the move, the process of opening up as such, that is important for the emotional appeal of such scenes, rather than the simple occupation of non-diegetic space. In all these cases, however, this technique leads to the paradoxical situation of 1960's characters actually playing records that were obviously made in the eighties.

6. Results: Towards a theory of multimodality

In general, the findings regarding the use of sound in *Dirty Dancing* seem to support Kress and van Leeuwen's tentative characterisation of different modes as having different metafunctional configurations. Vision, they posit, creates a sharp distinction between its subject and its object, whereas sound is more immersive (Kress & van Leeuwen 2001: 189–190). It may even be possible to extend this claim by saying that, as a likely result of different metafunctional configurations, different modes are used for certain functions to varying degrees. However, these configurations and their "affordances", as Kress and van Leeuwen (2001) call their metafunctional properties, are also historically and culturally contingent. Thus, sound in Western culture could be described as less 'suited' to the ideational than to the interpersonal metafunction of communication. In other words, its ideational meaning potential would be less extensive than that of its interpersonal metafunction (Kress & van Leeuwen 2001: 189–193).

Returning briefly to the aim of providing a characterisation of the kinds of meaning expressed through specific modes or modal configurations, one can summarise the findings by saying that it was possible to identify clear patterns concerning the elements encoding class membership, as well as class conflict and its

resolution in *Dirty Dancing*. Of the modes available to the medium, the makers of *Dirty Dancing* have used a large number to represent class membership, that is to say at least in the visual field – remarkably, however, dialect and register do not feature as markers of class in *Dirty Dancing*. In terms of both salience and narrative development, it is the dirty dancing of the lower classes that is by far the most significant mode of signification. In fact, as far as the diegetic world of *Dirty Dancing* is concerned, their style of dancing is the principle form of showing class membership, the principal way, that is, of *being* of a particular class. Closely related to this, of course, is the music that accompanies the dirty dancing, which differs from the music accompanying the dancing of upper class characters in several respects – perhaps most importantly in that it is exclusively African-American in the case of the lower class, whereas it never is with upper class dancing.

A number of modes are furthermore combined to charge the body of the lower classes with sexual meanings. Through the movements of the dirty dancers, their dress code, as well as the music accompanying their dancing, the act of their dancing is highly sexualised. This is why Baby's first encounter with the dirty dancing of the lower classes is often perceived as a "terpsichorean deflowering",[14] why it prefigures her actual first sexual encounter with Johnny, and why this encounter is fittingly initiated through dancing as a form of erotic foreplay.

7. Conclusion

In its final scenes and through complex multimodal configurations, *Dirty Dancing* achieves a sense of closure by resolving (read *dissolving*) through dancing what it has previously constructed through dancing as a semiotic mode. In order to find a similar sense of closure alongside the film, let us highlight a few aspects of of the final dancing scene along the lines of a multimodal analysis as suggested in this chapter. Since the makers of the film have staged all conflicts over boundaries of class, as well as the boundaries themselves, as a matter of dancing styles throughout, it becomes possible for the characters to ultimately 'transcend' these without ever having had to address them explicitly.

The final song of the film, a reprise of *The Time of My Life* by Jennifer Warnes and Bill Medley, takes the form of a duet. In van Leeuwen's (1999) terms, its structure is that of initiator – reactor (male – female), and contains a movement from segregation to unison. The song is furthermore one of the 1980's tracks of the film and also makes the move from diegetic to non-diegetic sound. In short, it has nearly all the features that van Leeuwen (1999) identifies as creating emotional appeal, as well as the movement characteristic of *Dirty Dancing* which opens up the diegetic space to involve the time and space of the spectator.[15]

The dancing in the final sequence serves a similar function as its soundtrack. As opposed to the various training sequences in the film, the final dance is perfectly choreographed and highly stylised. It is furthermore edited to create the impression of a flowing perfection and effortless grace, something that Baby and Johnny have achieved in the course of the film. In the words of producer and scriptwriter Eleanor Bergstein, the final scenes were choreographed and edited in this way so that the audience could "be in those steps", because "it's that time of your life when you do something so incredibly better than you have ever done it ever before [...] what we want to say is you will forever do it" (*Dirty Dancing*: special DVD edition 2001).

As regards the dissolution of class boundaries, it is important to bear in mind that all of this takes place in a space that has opened up to include the non-diegetic. It is Johnny's jump from the elevated stage and into the (diegetic) audience space that initiates the breaching and ultimately the dissolution of class boundaries. Johnny leaves the separate space of the stage – the space which he has always been allowed to occupy while dancing – in order to violate the classed space of the upper-class audience. He then invites the other dance people to join him, and when together they manage to induce everyone in the upper class audience to join them in their dirty dancing, they facilitate, strictly in terms of dancing, a perfectly innocuous mingling of the classes.

Not surprisingly, the actual closure of the film is only made complete with the formation of the heterosexual couple. Having won his victory over the classed system of dancing, Johnny is finally reunited with Baby. It is important to note that when he joins in to *The Time of My Life* by silently mouthing the male voice part – telling Baby he has "had the time of my life" and that he "owes it all to her" – this takes place in the non-diegetic space of surround music that is part of the intended audience's world (of the 1980's) and not of the characters' world (of the 1960's).

Indeed, there seems to be a functional relation between film-specific and non-specific modes and the concepts of diegetic and non-diegetic space in film theory. As they are not, however, identical, it may prove necessary but also extremely fruitful to explore in detail both the theoretical and the practical relationship between these concepts in the future.

Notes

1. Affinities to specific forms of cultural studies include, among others, the work of Stuart Hall (1992) and Lawrence Grossberg (1997). For more detailed information on the project's affiliation and the web anthology associated with it, see also http://www.univie.ac.at/Anglistik/easyrider/welcome.htm

2. See, for instance, Kress and van Leeuwen (2001:4).

3. This, of course, is not to say that the medium film cannot or could not do otherwise. In fact, doing work in cultural studies entails an opposition to the kind of essentialism that claims that there is some inherent 'essence' to film which filmmakers should strive to be true to.

4. Although colour is no doubt a significant resource in meaning-making as well, I have omitted it in my discussion here, as its status as mode (or system) is still controversial.

5. A good example of such transposition is the proliferation of filmic techniques in comics. Although the two media initially developed separately, comic artists soon began to take over certain cinematic 'devices' and techniques, such as the establishing shot, camera angles, or the shot-counter-shot technique used in dialogue. For more on the influence of filmic conventions on comics, see Roger Sabin (1993).

6. Christian Metz (1974), in his version of a map of what he called the "codes" of film, depicted specific and non-specific codes as concentric circles, with the more specific codes closer to the centre. While this does have certain advantages, it suggests that non-specific codes always include specific codes, something that Metz never intended.

7. See van Leeuwen (1999: 67–70, 150–154) for a discussion of voice quality and other criteria that distinguish language in song from speech.

8. These, it could be argued, are really systems of meaning, like colour, rather than modes. I would posit, however, that both *editing* and *camera movement* are well on their way of being established as modes, or already have been.

9. See the review of the movie by critic Roger Ebert, originally published by *The Sun Times* and now available at http://www.suntimes.com/ebert/ebert_reviews/1987/08/248895.html

10. At present, the terms 'distribution', 'configuration', 'combinatoire', or 'multimodal ensemble' are still used interchangeably (and somewhat arbitrarily). However, each has certain advantages as it emphasizes a specific aspect – the process of meaning-making, the structure, or the conjuncture of multiple modes – which is the reason for my use of more than one if not all of the terms in circulation.

11. The training or preparation sequence is, of course, an essential ingredient in every kind of sports film, and John Avildson's *Rocky* (1976) is a classic example.

12. These intrusions include *Wipeout* by The Surfaris, *Hungry Eyes* by Eric Carmen, *Overload* by Zappacosta, *Yes* by Mary Clayton, *She's Like the Wind* by Patrick Swayze, and the song that won an Oscar for *Dirty Dancing*, *The Time of my Life* by Jennifer Warnes and Bill Medley.

13. Remarkably, none of the tracks used to establish the 1960's 'sound' dates from later than 1963.

14. Review by Edwin Jahiel available at http://www.prairienet.org/ejahiel/dirtdanc.htm

15. See van Leeuwen (1999: 75–79) for an introduction to the terms used in his analysis of music. See also van Leeuwen (1999: 85–89) for a discussion of the musical form of the duet.

References

Butler, Judith (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

de Certeau, Michel (1985). "Practices of space." In M. Blonsky (Ed.), *On Signs: A Semiotics Reader* (pp. 122–145). Oxford: Blackwell.

de Certeau, Michel (1988). *The Practice of Everyday Life.* Berkeley: University of California Press.

Eber, Roger (1987). "Dirty Dancing." *Chigaco Sun-Times* 08/21/1987. http://www.suntimes.com/ebert/ebert_reviews/1987/08/248895.html

Gowaty, Patricia A. (1997). *Feminism and Evolutionary Biology: Boundaries, Intersections, and Frontiers.* New York: Chapman & Hall.

Grossberg, Lawrence (1997). *Bringing it All Back Home: Essays on Cultural Studies*. Durham: Duke University Press.

Hall, Stuart (1992). "Cultural studies and its theoretical legacies." In L. Grossberg, C. Nelson, & P. A. Treichler (Eds.), *Cultural Studies* (pp. 277–294). New York: Routledge.

Jahiel, Edwin (1988). "Movie reviews by Edwin Jahiel: Dirty Dancing." http://www.prairienet.org/ejahiel/ dirtdanc.htm

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge.

Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Edward Arnold.

van Leeuwen, Theo (1999). *Speech, Music, Sound*. New York: St. Martin Press.

Metz, Christian (1974). *Film Language: A Semiotics of the Cinema.* New York: Oxford University Press.

Sabin, Roger (1993). *Adult Comics: An Introduction*. London: Methuen.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Thibault, Paul (2000). "The multimodal transcription of a television advertisement: theory and practice." In A. Baldry (Ed.), *Multimodality and Multimediality in the distance Learning Age* (pp. 311–385). Campobasso: Palladino Editore.

Williams, Raymond (1961). *The Long Revolution*. New York: Columbia University Press.

Chapter 8

Multimodal text analysis and subtitling

Christopher Taylor
University of Trieste, Italy

For some years now this author has collaborated with colleagues at the universities of Pavia, Padua and Venice in research projects concerning multimodal text analysis and the creation of multimodal corpora, e.g. Taylor (2000). In particular the work of Thibault (2000) and Baldry (2000) has inspired work on the harnessing of multimodal transcriptions to the task of translating screen texts for interlingual subtitles. Specific translation (and other) strategies are required in subtitling film scripts and the like, and the multimodal transcription provides a scientific basis for formulating some of those strategies which are involved in the particularly controversial process of text condensation. A whole series of different multimodal text types have been analysed (feature films, soap operas, advertisements, cartoons, documentaries, etc.) and this chapter will illustrate the methodology underlying and the results obtained from a number of those analyses.

1. Introduction

The purpose of this chapter is to trace the search for useful translation strategies for subtitlers of video material. The traditional theories relating to translation in general are of assistance, in that subtitling is, after all, just one example of applied translation, though the specificities attached to this genre suggest that some particular strategies or approaches might be usefully sought. The chapter will first run briefly through the works and ideas of some of the major figures in the history of translation studies before investigating more recent contributions with more specific multimodal interests.

2. Translation studies

From the dawn of translation theory in the first century B.C., when Cicero posited the dichotomy of 'sense' versus 'word' (*ad sensum/ad verbum*) (cf. George Steiner 1975), the debate basically revolved around the following question – the spirit or

the letter, the matter or the manner? The more authoritative a text was considered, e.g. the Bible, the more 'literal' the translation should be. This has led in the past to self-doubt. Cicero himself declared "If I render word for word, the results will sound uncouth, and if compelled by necessity I alter anything in the order of wording, I shall seem to have departed from the function of a translator" (in Nida 1964: 13).

Centuries later, Dryden, who translated Horace, expressed his scepticism over the feasibility of being able to translate literally and well at the same time with: "Tis much like dancing on ropes with fetter'd legs!" (in Nida 1964: 18). Even in the twentieth century there has been no let up in the sense/word debate. Radically different positions have been established ranging from Walter Benjamin's (1968) "The sentence is a wall blocking out the language of the original, while word for word translation is the arcade" to Koller (1972) who propounded the idea of the equivalent-effect principle of translation, challenging the authority of the word or structure.

Accepting that taking a polarised position was counter-productive, theorists such as Eugene Nida (1964, 1996) made important distinctions between source-language oriented and target-language oriented translations and produced the concepts of *formal* and *dynamic equivalence.* His dynamic translation worked at the surface level in making a target text understandable to new audiences. Juliane House (1977) maintained the distinction between 'faithful' and 'free' by using the terms 'overt' and 'covert' in the sense that a more faithful version of a source text would be (overtly) recognisable as a translation, while a freer version would seem more like an original text in the target language (thus covert), and pointing out that it is the translator's responsibility to choose where to pitch the translation along the overt-covert axis. Peter Newmark (1982) gave us *semantic* and *communicative* translation, continuing to emphasise that different text types required different approaches, again harking back to the 'authority' of the text. It would seem that an authoritative text is one whose status or importance is such that "the manner is as important as the matter", requiring the translator "to empathise with the writer" (Newmark 1991: 109), while texts of lesser authority can be dealt with more freely.

Towards the end of the 19th century, when it was decided in some quarters that language was the product of culture, an enormous leap in logic concluded that translation was actually impossible. We have already seen how scholars such as Walter Benjamin held the view that translation should at least be as literal as possible. So, who knows what Benjamin would have made of the following translation (Example 1) from the film 'Kramer versus Kramer' (Benton 1979)? A divorced father, played by Dustin Hoffman, is explaining to his young son how lucky he is these days to have so many things that his father's generation did not possess:

(1) We didn't have the Mets, but we had the Brooklyn Dodgers, we had Polo Grounds and Ebbets Field.

The 'translation' of this line in the dubbed version of the film is as in (2):

(2) Non avevamo i motorini, ma avevamo i monopattini, le pistole a schizzo e non a laser.
We didn't have scooters, we had roller-skates, we had water pistols, not laser guns.

The justification for this departure from the original is that a literal translation would mean little or nothing to the target audience, and the meaning, a constant and universal parental lament, is maintained. The translation in this case is dynamic, covert and communicative, and the question now arises as to what extent this approach is valid within the film medium in general, if that medium is to be conceived of as in any way creative in the same light as literary works.

3. Creative texts

According to writer/translator Tim Parks (1997), creative literature 'declares its distance' from all other works, though he concedes that its translation is eminently possible in the right hands. He gives plenty of examples of what happens when the wrong hands are at work, but the implication is always that the target language has the necessary resources if enough effort is put into deploying them. One of Park's (1997: 13) examples taken from D. H. Lawrence's 'Women in Love' (3a) and one of its Italian translations (Lawrence 1989, transl. by dell'Orto) in (3b), and its back-translation in (3c) is illuminating:

(3) a. In a few minutes the train was running through the disgrace of outspread suburbia.
b. Di lì a qualche minuto il treno percorreva gli squallidi sobborghi della città.
c. *A few minutes from then the train was running through the squalid suburbs of the city.*

In terms of the traditional debate, this translation is inadequate because the sense is translated but the words are not. But this does not mean that a literal translation of the words is necessarily called for (*la disgrazia di una periferia distesa*?), rather that the translator has not recognised 'the distance'. What is required, Parks would argue, is some form of 'equivalent effect' declaring a distance in Italian. As far as screen translation is concerned, the concepts of both 'declared distance' and 'equivalent effect' are of interest. But the various semiotic modalities at play enable these effects to be implemented by more than just words. The manipulative

use of the camera, and the inventive use of light, colour, sound, etc. can help a film declare its distance. Think of the power relations established by the imaginative filming techniques in *Citizen Kane* (Welles 1941) already in the 1940s, the first innovatory use of the 'Steadicam' by Stanley Kubrick in *The Shining* (Kubrick 1980), and all the special effects now adopted by the industry. These manipulations are then understood, ignored, accepted, praised, criticised, or rejected by the viewer. The visual is a place where meanings are created and contested (Mirzoeff 1999: 6), and the viewer is a decisive element in the process.

In any film of *Women in Love*, the scene of the entry into London would be in the hands of the director. The 'translation' could then be criticised in similar terms to the written translation – does it just show a train running through the dismal outskirts or does it pan across a desolate urban landscape, in black and white, accompanied by dreary music, etc.? The viewer would then construe these 'meanings' in the way the director intended, or in some other way. However, these considerations are at a stage prior to the intervention of the subtitler, who is generally required to work on a video text that has already been made, and whose initial interpretative framework is already in place.

The question of 'equivalent effect' deserves more attention, though it too applies to all translation genres. Many modern translation scholars have made it clear they are dealing with 'text' (and 'text in context') rather than with 'words' or authors, and simultaneously there has been a move from the abstract to the specific: What are texts doing on the surface? How can their 'effect' be transferred? For example, Newmark's (1982: 39) communicative translations are those that create entire texts that "produce an effect as close as possible to that obtained in the reader of the original", a concept that pre-dated Newmark (Cauer 1896; Koller 1972) and has been oft repeated in different words by many other writers on the subject. The achieving of equivalent effect, be it to amuse, to frighten, to persuade, or whatever, is not always easy but examples abound to show what can be done. Sergio Jacquier's ingenious solution (see Galassi 1994: 62) to a Groucho Marx line will serve to show how all the components of a multimodal text can be used to create 'equivalent effect': a scene in *Horse Feathers* (Mcleod 1932) has Groucho desperate to conclude a deal and seal a document. "Get me a seal!" he yells at Harpo: Harpo obliges by bringing him an exemplar of the large sea creature known as a seal. The audience laugh as they appreciate the play on words and at the same time see the ridiculous result of the misunderstanding. They are also reminded of Harpo's stupidity, underlined as usual by his dumb expression and clown-like garb. Jacquier's Groucho makes no mention of a seal/sigillo but, consonant with his movements and gestures, he says "Focalizziamo" (Let's focus on this). Harpo, of course, still arrives with the seal, the word for which in Italian is 'foca'.

Returning to the general field of translation studies, a number of new insights into the process of translation have emerged in recent years, and they respond to the need to break out of the tight theoretical mould and address such questions as: 'What is the purpose of the translation?, What is the nature of the readership?, and What type of text are we dealing with?'. No single theory has provided all the answers, hence the move has been towards more broad-based approaches such as Susan Bassnett's (1980) 'general principles', Albrecht Neubert's (1995) 'models of translation', or Mary Snell Hornby's (1988) 'integrated approach', which is also reflected in a move towards the viewing of translation as a cross-cultural event.

To sum up the argument so far, all the above-mentioned aspects of translation theory are of interest, but where specifically do they take us in our search for strategies for dealing with film, in that area of our discipline which Sergio Viaggio (1992: 27) has described as "the least studied of all the branches of translation."

4. Subtitling

Subtitling indeed constitutes a relatively new text type and specifically addressed by hardly any of the above theories, schools, etc. It is an example of *intersemiotic translation*. The term is Jakobson's (1966) and refers to the fact that the meaning created in one modality (e.g. the visual) may be translated in another modality (e.g. the written language) or even vice versa in this digitally manipulative age, but it also simply means that the source and target texts consist of a number of interacting semiotic modalities. As a sub-group of screen translation texts, it has two essential purposes, entertainment and didactics, while containing a number of sub-sub-groups within these two broad areas. These two areas involve largely different audiences and approaches, as explained later, and these audiences themselves can be usefully categorised in terms of such factors as age, geographical provenance (e.g. subtitling or dubbing cultures, which in Europe conform to a basically North-South divide, though the principal reasons are financial rather than geographical – the countries with larger populations can afford to dub), and most importantly the level of knowledge of the source language. Furthermore, the subject material of screen texts varies as much as in non-film genres and has a major influence on strategies adopted – documentaries require different treatment to cartoon films for children.

But the very specificity of subtitling as a sub-genre of translation is important and is neatly subsumed by Gottlieb in the following way:

> In the context of translation, and expressed in general and rather technical terms, subtitling consists in the rendering in a different language of verbal messages in

> filmic media in the shape of one or more lines of written text presented on the screen in sync with the original verbal message. (Gottlieb 2000: 14)

He goes on to compare subtitling with other sub-types of translation, showing that it is the only genre that is a "*prepared* communication using *written* language acting as an *additive* and *synchronous* semiotic channel as part of a *transient* and *polysemiotic* text" (Gottlieb 2000: 15). Dubbing, for example, is not additive: it replaces the original soundtrack while subtitling provides an extra channel. It is also, of course, not written, at least at the time it is instantiated. Simultaneous interpreting, which has some aspects in common with subtitling, though it is actually less synchronous, is not (usually) prepared, not written, not additive, and not polysemiotic. Standard written translation is neither additive, nor synchronous, nor transient, nor polysemiotic. Thus it would seem clear that strategies for subtitling must differ from those for other forms of translation, particularly due to the fact that subtitled discourse *adds* something to the source text, and the fact that it is polysemiotic (or multimodal).

In the search for suitable strategic ways of thinking, firstly among the established theoretical frameworks, and looking within the world of linguistics, Malone's (1988) translation strategies and Halliday's (1994) functional grammar seem good places to start. Malone (1988) envisages four pairs of translating devices: equation/substitution, divergence/convergence, amplification/reduction, diffusion/condensation and one lone self-reciprocating element, re-ordering. It is here that synchronic linguistics meets translation studies, as Malone's approach offers structural and lexico-grammatical answers to a wide range of translation situations. The terms are reasonably self-explanatory, referring to translation strategies that operate along the cline between the two poles of 'total translation' in both semantic and lexico-grammatical terms and 'total freedom', whether this entails substituting lexis, expanding or reducing the original content, or in any way manipulating the source text to achieve the required effect. Then it is systemic functional linguistics that supplies the important macro-element required to see whether the text is working as a social semiotic, and to ensure the integration of interpersonal and textual considerations along with the ideational layer. Important work by Kovacic (1998) and Mason (1989) on the interpersonal element has also proved invaluable.

Malone's work is partly reflected in Gottlieb's ten translation strategies, specifically designed for subtitling, as can be seen from the similarity in some of the terminology (Gottlieb 1992): (1) *Transfer*, (2), *Condensation*, (3) *Expansion*, (4) *Deletion*, (5) *Decimation*, (6) *Imitation*, (7) *Dislocation*, (8) *Transcription*, (9) *Paraphrase*, and (10) *Resignation*. The most important elements in this scheme can be subsumed in the tension that lies between 'transfer', which is the translating of all the semantic elements present in the source text, and as much of the lexico-

grammatical organisation as the target text can bear, and the various stages of text shortening which come under the headings 'Condensation', 'Decimation' and 'Deletion'. When a text is 'condensed', the semantic content remains intact while the lexico-grammatical structures for expressing that content are reduced, for example: *Would you mind stepping this way?* and *Entri.* The formal exhortation to enter (*Entri*) uses the 'Lei' form of the imperative in Italian, and equates with the polite and elaborate form of the English invitation. If, on the other hand, a text is 'decimated', then a part of the semantic content is sacrificed, as in (4):

(4) You will have heard on the news that all the passengers were killed.
Lo sai che tutti i passeggeri sono morti.
'You know all the passengers are dead.'

'Deletion' implies a total loss of information. It would seem that these last three strategies would be those most used by subtitlers. The other semiotic modalities providing meaning on the screen or through the soundtrack (music, sounds, etc.) should, it would be supposed, allow not only for some creative condensation, but also a reasonable amount of selective decimation, and even deletion where the meaning is carried by visual or other verbal vehicles.

Indeed, these processes can be seen at work in a BBC documentary concerning a herd of female elephants suffering in the heat of an African drought with a crippled calf to look after (*Echo of the Elephants*, directed by M. Colbeck 1972). The soundtrack to one scene, out in the savannah, begins with : *But feeding and drinking meant abandoning Echo and the new calf.* Decimating one semantic element as being superfluous (drinking is subsumed in the generic *mangiare*), the Italian subtitle becomes: *Ma per mangiare devono abbandonare Eco e il piccolo.* The next line is much longer: *Ella, with her ragged ears, who had taken over the leadership of the group, turned back to Echo and Enid.* Here, the subtitler has first opted for the deletion of the description of the 'ragged ears' on the grounds that they have already been mentioned before during the documentary and the ears are there to be seen, in close-up, on the screen. Then, with an exquisite piece of condensation, the finite subordinate clause is condensed to a noun group: (*Ella*) *il nuovo capogruppo*, ... While these examples tend to indicate that subtitlers consciously or subconsciously search for reducing strategies, it is also true that the few statistics available (e.g. those of Gottlieb (1994, 1997) and those resulting from investigations carried out in Trieste) show that a large percentage of titles are examples of pure transfer. This point will be returned to after a discussion of some of the other theoretical contributions in the field.

Sylfest Lomheim's (1999) rather simpler set of parameters in Figure 1 also concentrates on the reduction/expansion axis, but interposes the rather interesting and usefully labelled item of 'translation'. Lomheim defines this category as

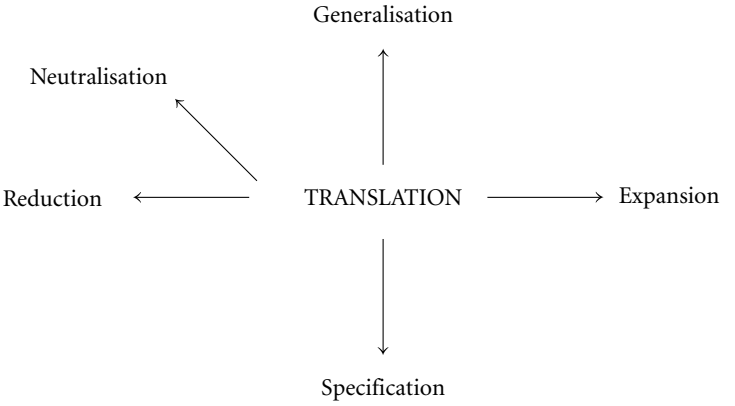


Figure 1. Lomheim's (1999) parameters

"a linguistic transfer that meets the normal requirements for equivalent translation". This is a useful distinction in that it calls into play previous definitions of 'condensation', 'expansion', etc. For example, the Italian *carta geografica* is not an expansion but simply the standard equivalent of the English *map,* and *map* is not a reduction of *carta geografica.* However, *la rappresentazione cartografica dell'Italia nella Galleria Belvedere*, translated as *the map in the Belvedere* is a case of reduction.

Lomheim also introduces the idea of 'generalisation', 'specification' and 'neutralisation', which also usually involve some degree of reduction or expansion, but the justification is for more than just spatial reasons. Generalising is necessary when culture-bound terms are used, as for instance in *'All she thinks about is going to Eastbourne' – 'Pensa solo al mare'*. Specification is required where there are, for example, lexical gaps in the source language, such as in the case of the potential ambiguity in Italian between a grand-daughter and the daughter of a brother or sister: *Vado da mia nipote – I'm going to my niece's.* Neutralisation comes into play when the translator wishes to or needs to attenuate (censorship? good taste? publisher's instructions?) or soften a source language utterance, e.g. *'What's the bloody matter! – 'Cosa c'è?'*.

5. The multimodal approach

Clearly all of the theoretical contributions referred to previously, and particularly those discussed in Section 4, are relevant also to screen translation, but what was still lacking until recently was an approach that really got to grips with the multimodal aspect itself. Hence the importance of the work carried out by Baldry (2000) and Thibault (2000), in the wake of the innovatory multimodal studies carried

out by O'Toole (1994), Kress and van Leeuwen (1996) and others. The *multimodal transcription* as devised by Thibault (2000) and Baldry (2000) provides an ideal tool for analysing the multimodal text in its entirety and drawing the relevant conclusions in terms of how meaning can be conveyed successfully by the various semiotic modalities in operation, and thus how dispensable or indispensable the verbal element is in different sets of circumstances. The multimodal transcription consists in the dividing of a video text into individual frames of, say, one second's duration. The frames are then arranged in sequence, vertically or horizontally, flanked by boxes arranged in columns or rows that systematically describe the various semiotic modalities of which the film is composed. In the pioneering article by Thibault (2000), a television advertising text for an Australian bank was transcribed and described in this way, providing a meticulous description in successive columns showing: (1) the time in seconds, (2) the visual frames themselves, (3) a description of the components of the visual image portrayed in terms of the camera position CP, the horizontal HP or vertical VP perspective, the visual focus VF, the virtual distance of the shot D, the visually salient items VS, the colours used CR, the coding orientation from natural to surreal CO, etc., (4) the kinesic action of the participants, (5) the complete soundtrack (dialogue, music, sounds, etc.) and (6) a metafunctional interpretation of how the film creates meaning as it unfolds over time, and a breakdown of the action into phases and sub-phases on the Gregory model (2002), within which various semiotic modalities are seen to function together as a set before giving way to a new set of modalities, following an identifiable 'transition' between the two phases. Figure 2 shows just the first two seconds of such a multimodal transcription of the afore-mentioned nature documentary *Echo of the Elephants*.

Observation of this kind of interpretation of how a multimodal text 'makes meaning' led the current author to consider using it as a basis for the judicious selection of the verbal element that needs to be maintained in utterances when subtitling a film. If the meaning, or a part of the meaning, of a section of multimodal film text is carried by semiotic modalities other than the verbal (visual clues, gesture, facial expression, dramatic music, surreal lighting effects, etc.), then a paring down of the verbal component can be justified, facilitating the various processes of condensation, decimation and deletion outlined above. Thus Thibault's multimodal transcription has been adapted for this purpose, on occasion by also fusing the visual image and kinesic action columns into one; the amount of detail provided by Thibault exceeds what is necessary for subtitling purposes. Then another column is added containing the subtitled version of the original soundtrack. Figure 3 shows the same two frames from the 'elephants' text, transcribed in this way.

Time	Visual frame SHOT 1	Visual image	Kinesic action	Sound-track	Metafunctional interpretation phases and subphases
1		CP: stationary HP. Frontal VP: median D: MLS VS: elephants VF: close CR: grey, green, blue CO: naturalistic	Animals grazing, moving their ears and tails, bird walking	Sounds of the savannah: birds tweeting; cicadas chirping	Phase 1 Subphase (a) Exp. Actors: elephants mill around. Int. Viewer positioned as observer of elephant's environment.
2		idem	A younger elephant enters on right and walks from right to left Tempo slow	"But feeding and drinking" . . . (pause)	Text. Hypothematic status introduces milieu

Figure 2. The first seconds of *Echo of the Elephants*

The example that will, however, be discussed here in a little more detail is taken from an episode of the BBC television comedy series 'Blackadder'. Humour is notoriously difficult to translate, especially apparently British humour, as it involves a large number of interweaving factors – word play, register shifts, timing, characterisation – to name but a few. However, this aspect is important but should not be exaggerated. What people of various cultures have in common is far greater than what separates them (see Nida 1996). There are often more radical extremes within cultures than between them. In this case, the same audience type in different cultures would appreciate this text, whereas there may be social or generational gaps within cultures which would determine very different reactions to it.

However, leaving aside the Buster Keaton era and the naughty slapstick of the likes of Benny Hill, much humour does rely heavily on the verbal, the delivery of the funny line. Consequently, in the subtitling of this episode of 'Blackadder', great care had to be taken in the subtitling in order to put the least burden on the spectator, who is expecting to be entertained and even to be occupied by paroxysms of laughter, and at the same time not miss any of the essential, and very effective, wording.

Time	Visual frame SHOT 1	Visual image + kinesic action	Soundtrack	Subtitle
1		CP: stationary HP: Frontal VP: median D: MLS VS: elephants VF: close CR: grey, green, blue CO: naturalistic Animals grazing, moving their ears and tails, bird walking	Sounds of the savannah: birds tweeting; cicadas chirping	
2		idem	"But feeding and drinking"... (pause)	"Ma per mangiare" ... (pause)

Figure 3. The transcription of first seconds of *Echo of the Elephants*

The episode in question, from the Elizabethan era series, features a highly implausible plot involving Lord Blackadder and his scatter-brained assistants, the dreadful Baldrick and Lord Percy, who have inadvertently executed the wrong man while temporarily in charge of the royal prison. The wife of the unfortunate victim, Lady Farrow, insists on seeing her husband, who she believes is still awaiting trial in the prison. Blackadder's scheme to extricate himself from this situation is to impersonate Lord Farrow at the meeting with his wife by wearing a bag over his head. Lord Percy has the job of explaining this to the unsuspecting lady.

In the first two 1-second frames (see Figure 4), Blackadder can be seen among his henchmen preparing to put the bag on his head. Lord Percy is nervously getting ready to meet Lady Farrow who is waiting outside. Blending an interpersonal interpretation onto this ideational description, the viewer sees the participants from the same conspiratorial level. Indeed, the viewer has a better perspective than any of the characters in that he/she has an unhindered view of all of them as they are arrayed on the screen. Blackadder is recognised as the boss – he has central position and the others occupy the margin, to use Kress and van Leeuwen's ter-

Time	Visual frame	Visual image + kinesic action	Soundtrack	Subtitle
1		Shot 1 CP: stationary/ HP: frontal/ VP: median/ D: MLS; VC: interior of the jail; Percy; Blackadder; Baldrick; Mr. Ploppy/ VS: the bag, exactly in the middle of the scene/ CO: artificial set; VF: distance: median; orientation: Blackadder and Baldrick gaze towards Percy **Kinesic action**: Blackadder orders Percy out by shouting to him/ Tempo: M	{RG} [ ] Blackadder: (**)*Go on,* (NA)*go on*// Pause/ Volume: f/ Tempo: F	Sbrigati! Sbrigati!
2		↓ **Kinesic action**: (Percy obeys and opens the door)/ Tempo: M	[opening of a door]	
3		Shot 2 CP: stationary/ HP: frontal/ VP: median/ D: MCS/ VC: outside the jail; Percy by the door; Lady Farrow before him; VS: Lady Farrow's proud posture/ CO: artificial set; VF: distance: close; orientation: Percy's gaze towards the floor (off-screen); Lady Farrow staring at him **Kinesic action**: [Percy walking past the door; Lady Farrow standing proudly before him]/ Tempo: M	[silence]	

Figure 4. The analysis of the 'Blackadder' episode

minology (1996:206): "For something to be presented as Centre means that it is presented as the nucleus of the information on which all the other elements are in some sense subservient." Their further comment that "the margins are these ancillary, dependent elements" fits Percy and Baldrick perfectly. Between the characters

themselves, Blackadder's withering gaze directed at Percy establishes the power structure and is a leitmotif for the whole series. The brainless Baldrick adopts the same stance, in slave-like imitation of his lord, another leitmotif of the whole series. Percy's nervous dithering tells us of the delicacy of his task.

From a textual point of view, the scene is the thematic element for the whole phase (see Gregory 2002), covering the fateful meeting, and marks a cohesive element with the third sub-phase when Percy re-enters the room. The set is an obvious mock-up of a prison, the costumes instantly recognisable as Elizabethan period, from Percy's fancy ruff to the rags that Baldrick wears as a member of the lowest social order, the colours and lack of colour playing an important role. This already prepares the audience for the incongruous actions that are to follow, which are at the heart of all humour. The audience subconsciously knows that humour is based on this premise and generally makes every effort to make sense of the text somehow, however bizarre it may be. They are helped by their intertextual knowledge of similar texts they have previously been exposed to. "Viewers always bring some idea of reality into the viewing process, testing the fiction for plausibility according to the worlds they know (both fictional and real), and adding their private associations to the specific sounds and images broadcast" (Spence 1995: 183). Even a patchy and scholastic knowledge of Elizabethan England prepares the viewer for the setting, knowledge of past BBC comedies, the style of Rowan Atkinson, and indeed past Blackadder series, prepare him or her for the kind of parody that will take place. The foreign audience may need a little more priming, especially preplanning in the form of prior publicity, articles in other media, etc., but the basic mechanisms come into play just the same. Otherwise, how could we account for the massive popularity of Brazilian 'telenovelas' on Russian television?

To turn now to the question of the translation, the only thing that is said in this brief scene is Blackadder's impatient injunction to Percy, *Go on! Go on!,* and this would thus not seem to tax the powers of the translator unduly. But conflicting pressures come to bear. In the interests of condensation, the obvious first step would be to remove the repetition, but Kovacic (1998) and Mason (1989) have both warned about the importance of interpersonal elements; here the repeated order is designed to express Blackadder's contempt for Percy and intense irritation at Percy's constant incompetence. He almost snarls the words. So we keep the repetition? At this point, the question of the audience arises. A minimum knowledge of the source language would equip any viewer with the necessary resources to interpret the text. And it is true that even those with no knowledge of the source language would still catch the aggressive intonation and the head movements expressing the feelings of the speaker. However, repetition of a word or short expression puts less pressure on the receptive capacities of a viewer than new material, and repeating the order would probably be the best option. This to-ing and

fro-ing between competing solutions reflects the thought processes of the translator as various options are considered, a process well illustrated by Krings' (1987) 'thinking aloud protocols', where translators were invited to record their thoughts on tape while they performed their task. But the problem still remains of what actual words to use. A literal translation into Italian would provide something like: *Avanti! Avanti!* But if the interpersonal elements are to be integrated, namely the contempt and the irritation, then a version incorporating a fairly colloquial verb plus the second person singular intimate pronoun (expressing the superior to inferior relationship), might be preferred: *Sbrigati! Sbrigati!*. The time taken to discuss this first minimum utterance is an indication of how much thought is required to translate a film for subtitles, but also shows how the multimodal transcription enables the translator to focus his/her efforts.

The next shot, consisting of nineteen frames (beginning at no. 3 in Figure 4), contains a little more dialogue. Percy exits from the prison cell to confront the waiting Lady Farrow. The slight differences in light and shadow from the darker dungeon to the relatively lit corridor are the cameraman's contribution. His photography is literally 'writing light'. This ability to 'write light' means that photography can create meaning not only by recording reality but by manipulating reality in the many ways adopted in filming. The 'film noir' was a good example of creating a jaundiced view of humanity. Intervention on the individual pixels in this digitalised age offers infinite possibilities of meaning-making. "The cameraman's point of view becomes ours" (Mirzoeff 1999: 103).

To return to the text, the ideational element consists in the fraught conversation in which Percy attempts to prepare the lady for the 'changed' appearance of her husband. The action is slow and the speech hesitant (hence the long shot) as Percy tries to gain time to allow Blackadder to hatch the plot. The viewer is again on the same level as the participants, enjoying the incongruous situation. The almost too perfect costumes maintain the pretence, and the constantly kept positions, suitably polarised on the left and the right, enable the viewer to easily interpret the respective stances: the proud yet distraught demeanour of Lady Farrow, who maintains her gaze steadfastly on the hapless Percy, and the embarrassment and totally inappropriate manner of the latter, who first attempts to avoid any eye contact.

The position of the two characters, repeated identically in phase 4, is a cohesive element, as is Percy's constant holding or fiddling with the door handle, the conduit between all the various scenes. The lady remains in the same position as a Given element, in that she represents sanity and the way things should be, while Percy's extravagant and fanciful inventions are the New element, and what provide the humour. Percy begins the dialogue slowly and hesitatingly, but is wrong-footed

by Lady Farrow, who shows an innocent belief that all will be well. Percy begins the difficult cover-up operation, see Example 5:

(5) Percy: Erm, sorry about the delay madam, er, as you know, you are about to meet your husband, whom you'll recognise on account of the fact that he has a bag over his head.
Lady Farrow: Why, I would know my darling anywhere.
Percy: Well, yes, there are a couple of things.

Percy's first speech would definitely seem to require some trimming, though whether this should consist of condensation, deletion, or decimation, in Gottlieb's terms, is the debatable point. Many of the elements are essential, both from an ideational and interpersonal, even textual point of view. The initial hesitant '*erm*' is a fundamental feature of spoken language, so often ignored by scriptwriters even though they should be striving to create texts that are 'written to be spoken as if not written'. Here it is used to express Percy's embarrassment as he begins his improbable explanation. On the one hand, it is not particularly language specific and can be heard quite clearly. But, on the other hand, it does function as theme for the first clause, albeit minor theme, and the speech is slow enough for the first element '*sorry about the delay, madam*' to appear alone on screen. However, it is known that the brain processes two lines appearing together quicker than two single lines appearing separately because of the onset of perception phenomenon. So, in the final analysis, after the usual to-ing and fro-ing, it might be better to split the whole speech into two double liners and make the '*erm*' the first sacrifice. The first line of the first two-liner will of course be '*sorry about the delay, madam*', which should be kept as it has a framing role and also because it contains one element that is in keeping with the period in question, namely the appellative *madam*. Such examples of period talk appear from time to time during every episode, precisely to provide the background against which the incongruous modern parlance can take full effect. The choice of translation for '*madam*' could be between a modern '*signora*', '*madama*' and a perhaps more appropriate expression, known and used in Italy in certain circumstances, also amusing circumstances. Hence '*scusate il ritardo,* ***milady***'.

There is then more hesitation and the discourse marker '*as you know*', which can be replaced in the subtitles with the punctuation device of three dots, which also serve the purpose of underlining the deliberate pause. The information '*you are about to meet your husband*' is of course central to the plot, but can survive a little decimation to '*ora vedrete vostro marito*', maintaining the second person plural form of address denoting the respect required and the old fashioned style.

The second two-liner begins with a fairly straight 'transfer' – '*lo riconoscerete*'/ '*whom you'll recognise*', though transforming the subordinate to a new main clause.

What follows is a key line in the humour 'on account of the fact that he has a bag over his head'. This line is delivered with a nicely measured pause, made more effective by the procrastinating conjunctive *'on account of the fact that'*. This time the procrastination can again be replaced by punctuation and by the insertion of a simple preposition, leading to the ludicrously incongruous '*dal sacco in testa*'. So Percy's speech would now be subtitled as: '*Scusate il ritardo, milady… ora vedrete vostro marito*' and *Lo riconoscerete… dal sacco in testa*'. Lady Farrow, seemingly unamazed by this news, declares *'Why, I would know my darling anywhere'*. The initial 'Why' is a melodramatic theatrical addition designed to augment the humour, but the stoic pose of the lady is probably sufficient to render this aspect for the foreign audience, and the depth of feeling present in her use of '*my darling*' comes out in the rather breathless delivery. The proud retort is thus condensed to *'Lo riconoscerei comunque'*.

The humour continues with Percy's next line '*Well, yes, there are a couple of other things*'. This line has to accompany Lady Farrow's, with the obligatory hyphens introducing the two parts of the dialogue. Although the two lines are not strictly connected as an adjacency pair, they should appear simultaneously as they form part of the same shot with the two characters on screen at the same time. In this case, the discourse markers and minor themes *'Well'* and *'yes'* are instrumental in Percy's strategy. The rest is not made explicit but the audience is party to the plot and enjoys the collusion. Thus the subtitles could also do a little conversation management *'Sì, però'*, to be followed by the more peremptory but equally (non) revealing *'C'è dell'altro'*, a classic case of condensation, in that no meaning has been deleted or decimated. Some indicative punctuation is also required here to contrast Percy's embarrassment with Lady Farrow's firm intent, although it is evident in the speech patterns. So these two lines would be: '*– Lo riconoscerei comunque.*' and '*– Sì … però … C'è dell'altro.*'.

The next shot is again short, consisting of only four frames in which we see a close-up of Lady Farrow, as she bows her head and lowers her eyes in order to deliver the carefully loaded line: *'I am prepared for the fact that he may have lost some weight'.* This is the first of six phases in which Lady Farrow is alone on screen and marks the beginning of an escalation in her emotive behaviour. From the quiet determination of this scene, she becomes ever more frustrated with the ridiculous Percy, and finally loses control in a hysterical crying fit. The scene therefore has an important thematic role and the rapport established with the audience at this stage is important. In spite of the fact that Lady Farrow avoids eye contact with the audience, they are still led to empathise with her predicament. The pathetic tone of voice and the uncalled for reasonableness of her attitude set the scene for the absurdity of Percy's response. As the subtitle will stand alone in the four seconds in which we see only Lady Farrow, there is no drastic need to curtail the line, and

thus the interpersonal and textual elements can be incorporated. Beginning with *'I am prepared for'*, this particular locution demonstrates the composure that the lady manages to maintain in the circumstances and requires an equivalent effect, perhaps '*Sarei orientata a credere che*'. The second part of the utterance needs a conditional to match the modal *may* and the slight paraphrase from weight to unit of weight, thus: '*Sarei orientata a credere . . . che potrebbe aver perso qualche chilo*'. The fact that the translation of this line, like so many others, could lead to heated debate on the appropriateness of one approach or another is indicative of how volatile the translation process is. Target texts are living things that can change form and function, style and meaning. We need all the help we can get to 'pin it down'.

The following three frames show Lord Percy in close-up as he delivers, with exquisite timing, one of the funniest lines in the whole series, given also the fact that Lord Farrow has been beheaded: '*Yes, and some height*'. Percy delivers this line while again fiddling with the door handle to his left (see above). He still needs to keep the lady waiting and provides this markedly tactless rejoinder. The phonological and semantic equivalences at play in the *'weight/height'* duo, can be partially rendered in Italian by converting to units of measurement, *'chilo/centimetro'*, which are also more acceptable stylistically. However, the key to creating the same effect on the target audience lies very much in the timing. The multimodal transcription can be very helpful here in ensuring that the 'spotting', that is the inserting of the title at the right moment, gains maximum effect. Again, suspension dots are useful. The whole sequence would now be subtitled as in (6):

(6) Sbrigati! Sbrigati!

Scusate il ritardo, milady. . .
. . .ora vedrete vostro marito'

Lo riconoscerete. . .
. . . dal sacco in testa.

– Lo riconoscerei comunque.
– Sì . . . però . . . C'è dell'altro.

Sarei orientata a credere. . .
che potrebbe aver perso qualche chilo

Sì . . . e qualche centimetro.

The episode continues in this vein and the translator is called upon to use every device available to create the equivalent effect. As there is some nuance, verbal or

otherwise, in practically every shot or phase, the multimodal transcription can be invaluable in focussing the translator on 'all' aspects of the text.

6. Conclusion

In conclusion, it is clear that all the vast store of knowledge that is now known as 'translation studies' may be brought to bear on every aspect of translation, including subtitling. The semantic/communicative distinction is important in assessing approaches to different types of multimodal text, as is the tendency towards either foreignisation or domestication. The concept of equivalence in its various guises, from Nida (1964), Wilss (1982), House (1977/1981), etc. through the variations on the theme of 'equivalent effect', to recent writers on screen translation (Gambier 1997, 1998; Galassi 1994), of course remains fundamental, but a major acknowledgement is due to all those who have worked in the field of multimodal texts and have devised various approaches to the analysis of those texts. In particular, the 'multimodal transcription' has provided screen translators with an extremely useful tool, one that has been exploited at length in Trieste in our work on subtitling different types of video text. In future, the advances demonstrated e.g. by Baldry (2000) and Thibault (2000) in the field of multimodal corpora and multimodal concordancing will be brought to bear on translation work. It is hoped that we soon can move to the field of subtitling film for didactic purposes in the promotion of minority languages, as well as hone our skills in the 'entertainment/information' section in the constant search for some failsafe strategies to make these texts instantly accessible to foreign audiences with the minimum effort expended.

References

Baldry, Anthony (2000). "English in a visual society: comparative and historical dimensions in multimodality and multimediality." In A. Baldry (Ed.), *Multimodality and Multimediality in the distance learning age* (pp. 41–90). Campobasso: Palladino Editore.

Bassnett, Susan (1980). *Translation Studies*. London: Routledge.

Benjamin, Walter (1968). "The task of the translator." *Delos, 2*, 76–99.

Cauer, Paul (1896). *Die Kunst des Übersetzens*. Berlin: Weidmann.

Galassi, Gianni G. (1994). "La norma traviata." In R. Baccolini, R. M. Bollettieri Bosinelli, & L. Gavioli (Eds.), *Il doppiaggio: trasposizioni linguistiche e culturali* (pp. 61–70). Bologna: Clueb.

Gambier, Yves (1997). *Language Transfer and Audiovisual Communication: A Bibliography*. Manchester: St. Jerome.

Gambier, Yves (1998). *Translating for the Media*. Manchester: St. Jerome.

Gottlieb, Henrik (1992). "Subtitling. A new university discipline." In C. Dollerup & A. Loddegaard (Eds.), *Teaching Translation and Interpreting. Training, Talent, Experience* (pp. 161–172). Amsterdam: John Benjamins.

Gottlieb, Henrik (1994). "Subtitling. People translating people." In C. Dollerup & A. Lindegaard (Eds.), *Teaching Translation and Interpreting 2. Insights, Aims, Visions* (pp. 261–274). Amsterdam: John Benjamins.

Gottlieb, Henrik (1997). "Quality revisited: the rendering of English idioms in Danish television subtitles vs. printed translations." In A. Trosborg (Ed.), *Text Typology and Translation* (pp. 309–338). Amsterdam: Benjamins.

Gottlieb, Henrik (2000). "Texts, translation and subtitling." In H. Gottlieb (Ed.), *Screen Translation* (pp. 1–40). Copenhagen: Center for Translation Studies.

Gregory, Michael (2002). "Phasal analysis within communication linguistics: two contrastive discourses." In P. Fries, M. Cummings, & D. Lockwood (Eds.), *Relations and Functions within and around Language*. London: Continuum.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold.

House, Juliane (1977/1981). *A Model for Translation Quality Assessment*. Tübingen: Gunter Narr.

Jakobson, Roman (1966). "On linguistic aspects of translation." In R. Brower (Ed.), *On Translation*. London: Oxford University Press.

Koller, Werner (1972). *Grundprobleme der Übersetzungstheorie*. Bern: Franke.

Kovacic, Irena (1998). "Six subtitlers – six subtitling texts." In L. Brookes (Ed.), *Unity in Diversity? Current Trends in Translation Studies* (pp. 75–82). Manchester: St. Jerome.

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge.

Krings, Hans-Peter (1987). "The use of introspective data in translation." In C. Faerch & G. Kasper (Eds.), *Introspection in Second Language Research* (pp. 159–176). Clevedon: Multilingual Matters.

Lawrence, D. H. (1997). *Women in Love*.

Lawrence, D. H. (1989). *Donne innamorate* (Translated by A. dell'Orto). Milan: Rizzoli.

Lomheim, Sylfest (1999). "The writing on the screen. Subtitling: A case study from Norwegian broadcasting (NRK)." In G. Anderman & M. Rogers (Eds.), *Word, Text, Translation: Liber Amicorum for Peter Newmark* (pp. 190–207). Clevedon: Multilingual Matters.

Malone, Joseph L. (1988). *The Science of Linguistics in the Art of Translation*. Albany: State University of New York Press.

Mason, I. (1989). "Speaker meaning and reader meaning: preserving coherence in screen translation." In R. Kölmer & J. Payne (Eds.), *Babel: The Cultural and Linguistic Barriers between Nations* (pp. 13–24). Aberdeen: Aberdeen University Press.

Mirzoeff, Nicholas (1999). *An Introduction to Visual Culture*. London: Routledge.

Neubert, Albrecht (1995). "Text and translation." *Rivista internazionale di tecnica della traduzione, 1*, 63–67.

Newmark, Peter (1982). *Approaches to Translation*. Oxford: Pergamon Press.

Newmark, Peter (1991). *About Translation*. Clevedon: Multilingual Matters.

Nida, Eugene (1964). *Towards a Science of Translation*. Leiden: E. J. Brill.

Nida, Eugene (1996). *The Sociolinguistics of Interlingual Communication*. Brussels: Editions du Hazard.

O'Toole, Michael (1994). *The Language of Displayed Art.* London: Pinter.

Parks, Tim (1997). *Tradurre l'inglese*. Milan: Bompiani.

Snell-Hornby, Mary (1988). *Translation Studies: an Integrated Approach*. Amsterdam: John Benjamins.

Spence, Louise (1995). "They killed off Marlena, but she's on another show now: Fantasy, reality and pleasure in watching daytime soap operas." In R. C. Allen (Ed.), *To Be Continued: Soap Operas around the World*. London: Routledge.

Steiner, George (1975). *After Babel*. London: Oxford University Press.

Taylor, Christopher (2000). "Text analysis and translation: An interactive, self-access computer application incorporating a functional approach." In A. Baldry (Ed.), *Multimodality and Multimediality in the Distance Learning Age* (pp. 295–310). Campobasso: Palladino Editore.

Thibault, Paul (2000). "The multimodal transcription of a television advertisement: theory and practice." In A. Baldry (Ed.), *Multimodality and Multimediality in the Distance Learning Age* (pp. 311–385). Campobasso: Palladino Editore.

Viaggio, Sergio (1992). "Contesting Peter Newmark." *Rivista internazionale di tecnica della traduzione, 0*, 27–58.

Wilss, Wolfram (1982). *The Science of Translation*. Tübingen: Gunter Narr Verlag.

Chapter 9

Multimodality in the translation of humour in comics

Klaus Kaindl
University of Vienna, Austria

In the few works on the translation of comics, discussion of the humorous dimension tends to be limited to verbal humour, i.e. playing with names and puns. The examples are almost always taken from 'linguistically demanding' comics such as *Astérix* or *Tintin*. This chapter sketches an approach which encompasses not only the verbal but also the non-verbal dimension of humour in the translation of comics. After a short discussion of the various forms of multimodal humour in comics and possible translation strategies, the chapter focuses on the semiotic complexity of playing on verbal and non-verbal signs. In order to provide a comprehensive account of humour, the chapter also explores visual comic techniques, e.g. pictorial intertextuality and visual allusions, and briefly discusses the comic potential of typography and onomatopoeia.

1. Introduction

The comic strip is a hybrid genre, whose analysis cannot be clearly assigned to any one academic discipline. This is due, among other things, to the techniques involved in designing comics, ranging from various linguistic elements such as text in speech bubbles, narrative texts, onomatopoeia and captions, to typographic elements such as typeface and type size, pictographic elements such as speedlines, ideograms such as stars, flowers etc., and pictorial representations of persons, objects and situations.[1] All these techniques play a part in conveying the meaning. The use of various modes in one genre makes it difficult for 'monomodal' disciplines to deal with such texts. Instead of analysing texts as a whole, they are segmented and become objects of study for various disciplines. For comics there are linguistics to speak about the language, graphic arts to deal with the pictures and communication studies to discuss the forms of publication and distribution.

When speaking of multimodal texts in this chapter, I mean texts in which various semiotic vehicles, e.g. language, image, sound, music etc., are used to convey meaning and to create a message: "Modes are semiotic resources which allow the

simultaneous realisation of discourses and types of (inter)action" (Kress & van Leeuwen 2001: 21). I shall take one specific aspect of communication, i.e. humour, and draw attention to its multimodal features. The aim is to outline the multimodal implications of humoristic practice in comics and the consequences for translation.

Although many kinds of text with multiple types of signs are dealt with in translation studies (especially film translation, advertising, theatre translation, etc.), the focus tends to be limited to their linguistic features. The interrelations between verbal and non-verbal signs, the various means of providing information by non-verbal elements and the transfer of non-verbal components in the process of translation are usually neglected. Thus, this chapter presents a framework for studying the translation of humour in comics as a multimodal practice. On the premise that translation is a cultural, rather than linguistic, phenomenon, I shall focus on the following questions:

– If language is only one part of the textual complex, what consequences does the multimodal composition of a text have for various translation strategies?
– How do word and image cooperate in the creation of meaning, and what role do these relations play in the transfer from one culture to another?
– If we do not translate languages but cultures, what is the role of the non-verbal dimension in translation: do we have to redefine the concept of translation in order to also include forms of transfer which do not involve language?

2. Multimodal humour and its translation

While studies of the general area of humour are mainly concerned with jokes, the study of comics tends to foreground other comic techniques. Jokes can generally be described as autonomous textual entities with a similar sequential organisation (i.e. introduction, text, reaction), which are structured so as to lead up to a punch line and are not necessarily dependent on contextual factors (cf. Attardo 1994: 296–311). Comics, in contrast, are narrative texts (whose plot need not be comical) which contain humorous elements but whose comic effect results from the overall narrative context. Rather than jokes, comics often work with techniques such as verbal and non-verbal puns, parody, allusion and intertextual reference. Their comic effect comes not from the punch line, which Kotthoff defines as a "clash between two perspectives", but from the "dual perspectivisation" of different contexts (Kotthoff 1996: 250).

In the existing literature on the translation of comics, only a limited number of the possible ways of creating comic effects have been discussed as being relevant to translation. Most of the studies deal with word-play in comics and focus primarily

Table 1. Translation techniques

	Source text: Semiotic type	Target text: Semiotic type	Source vs. target text: Humour technique		Source vs. target text: Language-picture relation	
A	Monomodal humour	Monomodal humour	similar	changed	similar	changed
B	Monomodal humour	Multimodal	similar	changed	changed	
C	Monomodal humour	No humour	–	–	–	–
D	No humour	Monomodal humour	changed		changed	
E	Multimodal humour	multimodal humour	similar	changed	similar	changed
F	Multimodal humour	Monomodal humour	similar	changed	changed	
G	Multimodal humour	No humour	–	–	–	–
H	No humour	Multimodal humour	changed		changed	

on the verbal element, ignoring, for the most part, the multimodal implications. The same applies to the pictorial depiction of situations, events, and persons. Although the polysemiotic constitution of comics is mentioned time and again, it is extremely rare to find it taken seriously in concrete analyses. If the comic element of this genre is to be analysed in all its variety and complexity, however, the narrow orientation of existing approaches must undoubtedly be overcome, and a broader framework be developed.

Based on the five picture-related categories, which the translator of humour in comics can be confronted with, and the respective degree of semiotic complexity, we can posit various types of translation solutions. These techniques (A to H) are presented in Table 1.

Monomodal humour refers not only to the linguistic dimension but also includes comic effects based exclusively on pictorial elements. The translation can have an impact on the semiotic composition of the sign-play, on the type of humour employed, and on the relation between text and picture underlying the comic effect. The choice of translation strategy is dependent on semiotic functional factors as well as pragmatic factors such as the cost of making changes to pictures, editorial policies and the intended readership. We may assume that particular categories of sign-play are associated with certain types of translation technique, but this hypothesis could be tested only in a more comprehensive study.

The present chapter therefore focuses on the systematic description and exemplification of the various categories of multimodal sign-play and techniques of visual humour.

3. From playing with words to playing with signs

Non-verbal elements in multimodal texts not only perform the function of illustrating the linguistic part of the text, but also play an integral role in the constitution of the meaning, whether through interaction with the linguistic elements or as an independent semiotic system.

If the visual elements in comics can also be put to use in creating word-play (or sign-play) effects, there is a need for a closer analysis of the possible relations between language and pictures. The semiotic complexity of a play on signs depends on the degree of integration of the signs involved and on the roles played by language and picture in the creation of the comic effect. An analysis reveals five distinct types of play on words and/or signs:

1. plays on words consisting basically of linguistic signs
2. plays on words reinforced by non-verbal signs
3. plays on signs that depend on a multimodal combination
4. non-verbal plays on signs reinforced by verbal signs
5. plays on signs consisting only of non-verbal elements

As general studies on visual puns have shown, non-verbal plays on signs function basically on the same principles as verbal puns; i.e. they are based on dichotomous relationships, such as that between form and content, and employ the polysemy of visual signs (cf. Lessard 1991). In order to recreate multimodal plays on signs in translation, it is essential to recognise the relations between the verbal and non-verbal signs and to analyse the function of the non-verbal elements. What factors need to be considered in translating a play on signs will depend to a great extent on what roles the respective pictorial and linguistic elements play in creating the effect.

3.1 Linguistic word-play

If the picture plays only a supporting role and the play on words is still understandable without relating it to the picture, then this relationship can be of subordinate importance in the translation as well. In that case, translation of humour in comics is not very different from translation of humour in non-pictorial texts (for the description of translation techniques for verbal puns, see Delabastita 1993: 33–39).

3.2 Plays on words reinforced by non-verbal signs

Plays on words reinforced by non-verbal signs are those in which the comic strip picture has a bearing on the linguistic play on words which reinforces, but is not integral to, its effect. An example of this type of word-play is a sequence from *Astérix le Gaulois* given in Figures 1 and 2.

The Romans have taken the Druid and Astérix captive and are determined to discover the secret of the magic potion. But instead of brewing the desired strength-giving mixture, the Druid concocts a hair restorer. Furious, the Roman commander demands an antidote. Astérix reacts to this in a number of idiomatic expressions which include lexemes with the semantic indication for 'hairy'. He comments on the Druid's unwillingness to brew an antidote with the French idioms "il a un poil dans la main" (literally: "he has a hair in his hand", figuratively:

Figure 1. Play on words reinforced by non-verbal sign in French

Figure 2. Play on words reinforced by non-verbal sign in German

"he does not feel like working") and "il a un cheveu sur la langue aussi" (literally: "he also has a hair on his tongue", figuratively: "he also lisps"). These remarks are accompanied by gestures which illustrate the literal meaning of the idioms: in the first panel Astérix is pointing to his hand and in the second, to his tongue. In this case, the comic effect derives from the situational context; the visual elements, i.e. the gestures, reinforce this effect, but are not integral to the constitution of the play on words. The German translation also employs a 'hairy' lexeme. This occurs in the left panel and prepares the reader, to a certain extent, for the play on words that then appears in the right panel, when Astérix says that the Druid has "Haare auf den Zähnen" (literally: "hair on his teeth", an idiom meaning: "he's a tough customer") and adds "und manchmal auf der Zunge" ("and sometimes on his tongue"). The latter is not a German idiom. In this way, one type of humour (two meaningful puns with a sarcastic, ridiculing effect) is replaced by another type of humour (one pun and one absurdity) but with a similar humorous effect of ridiculing. The German "und manchmal auf der Zunge" does not have the double pun meaning and is therefore absurd, and contrary to Grice's Cooperative Principle, which Astérix is supposed to respect vis-à-vis the Romans, who are thus not being taken seriously but are ridiculed (see Raskin 1985 and Attardo 1994 on the role of the four conversational maxims of Grice's Cooperative Principle in humour). The German version therefore preserves the polysemiotic nature of the picture-supported word-play (translation technique E) while at the same time changing the type of humour.

3.3 Plays on signs depending on a multimodal combination

In plays on signs that depend on a semiotic combination for their effect, interplay between verbal and non-verbal elements is integral to the comic effect. This type of interplay is illustrated by Figures 3 and 4 from *Tintin–Le trésor de Rackham le Rouge*.

Here the comic effect is dependent on the relation between the inscription in the picture and graphic signs. Captain Haddock has bumped his head against an advertisement pillar, an injury which is indicated graphically by stars. This creates a relation of identity with the literal – as opposed to the actual – meaning of the linguistic text on the pillar. The text is an advertisement for a daily newspaper and reads: "des informations qui frappent", an idiom which actually means "news which attracts attention", but the literal sense of which is "news which strikes". In the German translation, the content of the linguistic text on the pillar has been changed; it proclaims that advertising in the newspaper *Morgenpost* can be expected to bring "durchschlagenden Erfolg". This idiom means "total success", but the literal sense of "durchschlagen" is "to knock through" (from "schlagen" mean-

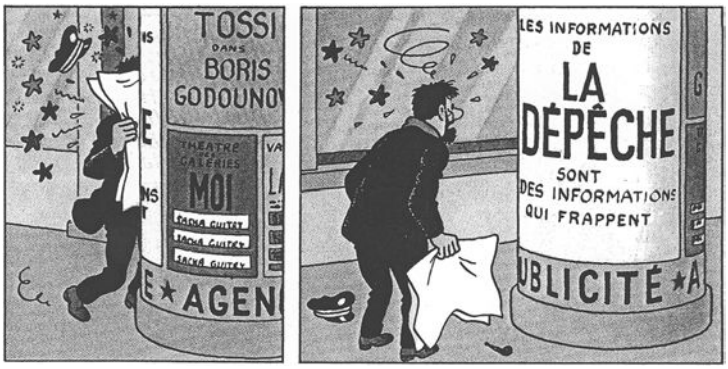


Figure 3. Play on signs depending on a multimodal combination in French

Figure 4. Play on signs depending on a multimodal combination in German

ing "to strike") so that, although the inscription is different, the play on signs and the underlying relation of identity between the graphic and linguistic elements have been retained. Once again, translation technique E was used, without changing the relationship between text and picture.

In principle, however, the underlying relationships between verbal and non-verbal elements may be changed in translation as well. An example of this can be found in the story *Le Bouclier d'Arvergne*, when Astérix and Obélix are forced to hide from the Romans overnight in a heap of coal, given in Figures 5 and 6.

In the French original, the play on signs is created by the contrast between the literal meaning of "passer une nuit blanche" (literally: "spend a white night", figuratively: "spend a sleepless night") and the picture showing the black heap of coal. In the German translation, the play on signs is realised through a relationship of identity between the literal meaning of the idiom "Ich sehe da in jeder Richtung

Figure 5. Play on signs depending on a multimodal combination based on contrast in French

Figure 6. Play on signs depending on a multimodal combination based on identity in German

schwarz" (literally: "I see black in every direction", figuratively: "I am pessimistic") and the black heap of coal.

Translation strategy G (from multimodal humour to no humour) can be exemplified by the following passage from *Astérix et les Goths* in Figures 7 and 8.

The chief of the Goths has been put in chains by his interpreter. Astérix lets him have the magic potion, and the chief manages to break free of his chains, upon which Astérix remarks: "il est dechainé" (literally: he is free from chains, figuratively: "he is beside himself with rage"). In the German translation, published in 1965 under the title "*Siggi und die Ostgoten*", there is only a comment on the picture ("wie die Fesseln fliegen", roughly: "how the chains are flying"). The comic effect resulting from the identity relation between the literal meaning of an idiomatic phrase and the scene in the picture is lost.

Figure 7. Play on signs establishing multimodal humour in French

Figure 8. No play on signs leading to loss of multimodal humour in German


Et ils n'ont pas tort, car Satiric, beau-frère de Liric, sous prétexte d'un repas de famille, tend un guet-apens, dont Liric est victime. C'est au cours de cette bataille qu'est née l'expression célèbre : « Ah, la famille !... »

Figure 9. No play on signs, no multimodal humour in French

The opposite case, i.e. no play on signs in the original but multimodal wordplay in the translation (translation technique H), can be found in Figures 9 and 10 from the second translation of *Astérix et les Goths*, dating from 1971.

Figure 10. Play on signs establishing multimodal humour in German

The French original merely mentions that "an ambush was laid" during a family dinner ("il tend un guet-apens"), whereas the translation creates an identity relation between the picture and the literal meaning of the German idiom "in die Pfanne hauen" (literally: "to throw somebody into the pan", figuratively: "to inflict a crushing defeat").

3.4 Non-verbal plays on signs

In addition to verbal word-play, non-verbal sign-play is also to be found in comics, with or without verbal reinforcement. A transfer to another cultural context becomes particularly problematic when the sign-play involves visual representations of objects that are specific to a particular culture. This type of problem is illustrated by a strip from *Krazy Kat* by George Herriman, shown in Figure 11.

Figure 11 depicts a blind pig, which, according to the story, is only pretending to be blind in order to scrounge money from others. In the U.S.A. at the time of prohibition, the expression "blind pig" meant a place where people could illegally consume alcohol. The comic effect derives from the portrayal of the literal sense of the expression in a pictorial code, which is also verbally reinforced by a placard with the word "blind" printed on it. In the German translation, the picture was retained; the translator merely added a note at the end of the volume explaining the play on signs for the German-speaking reader. While there is some merit in the explanation, the translator failed to recreate the comic effect stemming from the polysemy of the pictorial representation. As in Figures 7 and 8, this is a case which involves the deletion of humour, in this case, monomodal (pictorial) humour (translation technique C). In contrast to Figure 8, readers gain at least some understanding of the humour of the scene, thanks to the translator's explanatory endnote.

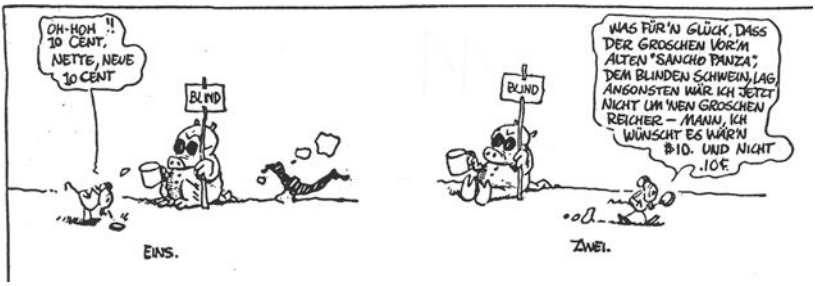


Figure 11. Culture-specific non-verbal play on signs

4. Comic pictures

Pictorial elements have generally been regarded as irrelevant to translation. This neglect stems from the fact that scholars of translation, in the early developmental stages of the field, focused on the linguistic dimension of their object of study. Multimodal texts, such as films, advertising, theatre, or comics came to attract increasing attention only in the 1980s, as translation studies became established as a discipline in its own right. Even so, to this day some translation scholars seem to believe in the existence of a universally comprehensible pictorial language. Rabadán (1991: 154), for instance, states that pictures represent a universal code which is therefore not subject to translation. The belief that there is a kind of visual Esperanto is rooted in the assumption that, unlike arbitrary linguistic signs, pictures exhibit a direct similarity to the reality they represent. However, Eco (1972: 202), for one, has demonstrated that pictures have a code which is governed by conventions, and these conventions may be shaped by cultural constraints. This also means that the visual representation of objects, gestures, facial expressions, etc. can be interpreted correctly only if the significance of these elements has been defined in the particular culture (cf. Eco 1987: 65).

Comics are very strongly governed by conventions (cf. Kloepfer 1976: 43), both in terms of the positioning of individual panels, i.e. the macro-structure, and in terms of the pictorial representation on the micro-structural level, i.e. within the panels. Moreover, the depiction of culture-specific objects or behaviours can lead to comprehension problems in translation. This also affects the rendition of comic elements, as will be shown by a few examples below.

4.1 Culture-specific pictorial elements

Discourse-analytical studies of humour have shown that non-verbal elements like facial expressions and gestures often play a significant role in the perception of

Figure 12. Loss of culture-specific multimodal humour

a spoken utterance as humorous or funny. This interdependence between verbal and non-verbal components is also significant in the way behaviour is depicted in comics in which actions and dialogues are closely intertwined with the visual representation of the characters. A good example for this can be found in the story "*L'homme aux 7 douleurs*", from *Les frustrés* by Claire Brétecher, given in Figure 12.

In Figure 12, a hypochondriac keeps pestering his girlfriend with his imagined ailments. The comic effect in this case results from the tension between the woman's routinely unimpressed response to the fears of the hypochondriac, and the latter's (mis)perception of this response as showing genuine concern for his complaints. This is conveyed both verbally and by gestures. The latter include the swearing of an oath, with which the woman unflinchingly assures the man that she would not hesitate to tear out with her own hands any tubes to which he might be connected for life support. Unlike in France, where the swearing of an oath is indicated non-verbally by an extended forearm and by spitting on the floor, the corresponding gesture in the German-speaking area would be a raised hand, the index and middle fingers of which have previously been put to one's tongue. Since German readers will be unfamiliar with the French gesture and will not recognise it as representing the swearing of an oath, the comic effect created by the woman's non-verbal behaviour will in part be lost. Hence, Figure 12 shows a reduction of multimodally mediated humour to monomodal (verbal) humour (translation technique F).

Nevertheless, the pictorial adaptation of such culture-specific elements is not always the ideal solution for maintaining the comic effect. For instance, in Figure 13 from the German version of *Peanuts*, by Charles Schulz, as published in the magazine *Stern*, the oval American football is replaced by the round (soccer) ball more commonly found in Europe.

The football is part of a running gag: Lucy will hold the ball, and whenever Charlie Brown tries to kick it, she pulls it away, sending Charlie to take a dive onto

Figure 13. Culture-specific change in the type of humour

the ground. In the American version, the action of holding the ball makes perfect sense, since the oval ball used in American football needs to be set up in this way for the kick. In the German version with a round ball used for soccer, whose rules do not permit players to touch the ball, the act of holding the ball appears unmotivated or to serve no other purpose than pulling it away, thus rendering the gag much less effective. Though it is true that the monomodal (pictorial) comic effect is maintained in German (translation technique A), we find a change in the type of humour. Whereas the original uses slapstick-like humour, with many repeated instances of a violation of expectable behaviour, Lucy's behaviour in the German version is unusual from the start; the act of holding the ball has no counterpart in reality, which renders the comic somewhat absurd.

4.2 Pictorial intertextuality

The creation of comic effects by visual means also relies heavily on intertextual relations, which, according to Genette (1982), include quotations and plagiarisms as well as textual allusions and, in particular, parodies. Following Kotthoff (1996:264), a parody is defined here as the imitation and functional transformation of source-text structures. The understanding of parody is culture-bound. Apart from verbal parodies, comics often use pictorial parodies, mainly of popular images or famous paintings. For instance, in *Astérix en Corse*, in Figure 14, Goscinny and Uderzo make a travesty of historical paintings of the Battle of Austerlitz in 1805, such as those by Vernet, Gérard, and de Roehn, and of the so-called *images d'Epinal* fashioned after these models.

Central to these is the rising sun, which gave Napoleon's *Grande Armée* a clear view of the enemy and thus helped it win the battle. The pictorial parody, which transplants the events to Corsica, is complemented by a verbal allusion as well as a phonological pun, both of which give linguistic expression to the link with Austerlitz: The leader of the Corsican rebels is called "Osterlix", which has practically the same pronunciation in French as Austerlitz. In addition, the guide in the bottom right panel speaks of the *sommeil d'Osterlix*, which is a playful transforma-

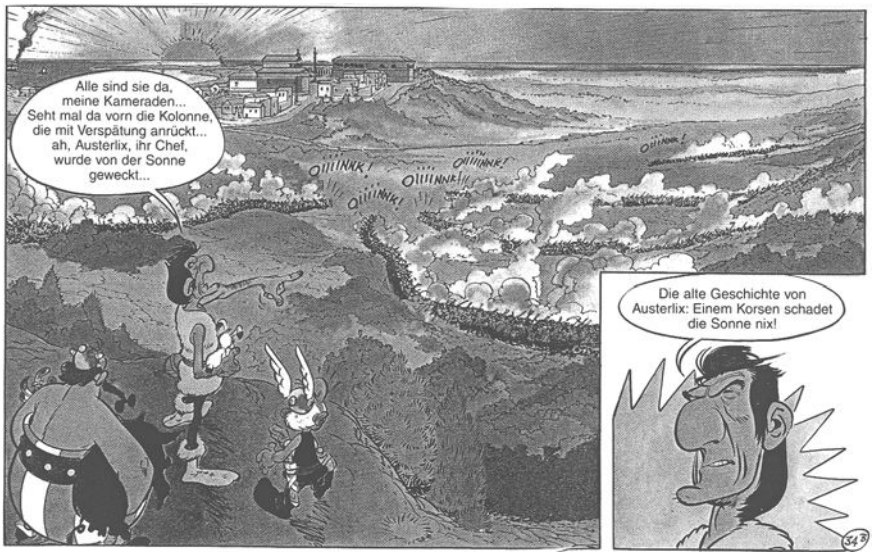


Figure 14. Reduction of multimodal humour to monomodal humour due to lack of culture-specific background knowledge

tion of *soleil d'Austerlitz*, as represented graphically by a schematic solar shape in the background. In the translations, the parodic effect of this sequence is lost, for lack of cultural background knowledge of the pictorial traditions for representing this event, and the multimodal humour is reduced to a monomodal word-play (translation technique F).

4.3 Pictorial allusions

A similar example is the parodical allusion to a famous painting by Géricault in the story "*Coke en stock*" by Hergé in Figures 15 and 16.

Tintin, his companion Haddock and a pilot named Szut have been shipwrecked and are seen floating on a raft. When they catch sight of a steamer which might rescue them, Haddock jumps with joy, breaks the raft and falls through it into the water. When he re-emerges with a jellyfish on his head, Tintin comments: "Vous voulez donc à tout prix que ce soit réellement le Radeau de la Méduse?" The latter makes reference to Géricault's painting "Radeau de la Méduse", which depicts the survivors of the shipwrecked frigate Méduse on a raft. The parodical reference is highlighted by an additional pictorial pun which relates the jellyfish (in French: *méduse*) on Haddock's head to the ship named after the Medusa of Greek mythology. The German translation follows the French original neither in the pictorially based pun nor in the parodical allusion to Géricault's painting. In-

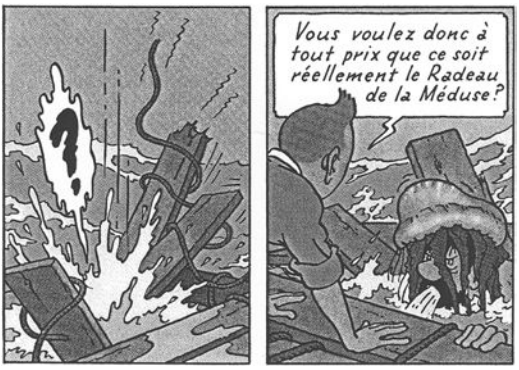


Figure 15. Multimodal humour based on pictorial allusion and pun in French

Figure 16. Deletion of humour based on culture-specific allusions in German

stead, Tintin asks Haddock, as he re-emerges spouting sea-water, whether he is still thirsty. The multimodal humour has been deleted (translation technique G).

Another type of allusion which is often found in comics involves elements of the visual culture. Whereas Genette limits the notion of allusions to textual relations, Wilss (1989: 45) conceives it more broadly as covering also phenomena which refer to culture-specific knowledge. Such presuppositions are found especially in comics in which the story is closely linked with the social, political, ideological, etc. cultural background. A good example for this is the character of the prime minister in the French comic *Monsieur le Ministre* in Figure 17.

The main character, which is drawn with the physical features typical of newspaper caricatures, alludes to Jacques Chirac, who served as prime minister under socialist President François Mitterand in the 1980s and made several failed attempts to win the presidency before finally succeeding in 1995. The depicted prime minister's acerbic facial expression plays on the French readers' knowledge that

Figure 17. Loss of humour due to pictorial allusion to particular political figure

Figure 18. Loss of humour due to culture-specific, different allusions to Little Red Riding Hood

Chirac would rather have been President. In the German translation, the lack of this background knowledge blocks the comic effect intended by this allusion (translation technique C).

Even where the allusion is to phenomena which are shared across cultures, like certain fairy-tales for example, translation problems can arise where there are differences in the visual images used. This is the case in Figure 18 from the story *Coke en stock* by Hergé.

When Tintin and Haddock return home, Tintin's dog Milou greets them with a pitiful howl. He has been dressed up in a strange costume – a red hat and a

pink cloak – by Abdallah, the son of an emir, who is friends with Tintin and Haddock and is currently staying at their house. This costume, which Milou evidently considers a disgrace, alludes to Charles Perrault's *Le Petit Chaperon Rouge* (Little Red Riding Hood), whose main character is traditionally depicted in such garb in French books of fairy-tales. The comic effect arises from the contrast between the naïve character of Little Red Riding Hood and the clever detective dog Milou, to whom Hergé attributes typically human behaviour (as in the present case, where the dog is embarrassed to tears by the ridiculous costume). In the German-speaking area, Little Red Riding Hood is usually depicted differently, i.e. with a red cap and a coat, which makes it difficult for the reader to appreciate the humorous contrast between the fairy-tale and the comic character (translation technique C). This example confirms Eco's (1987) claim that pictures of objects, in this case clothes, will be interpreted or 'read' only within a functional context, which may in turn be culture-specific.

5. Comic aspects of typography

The visual repertoire used in comics includes not only iconic signs, such as images of characters and objects, but also graphic and typographic elements which help create a comic effect. It is particularly in comics that typography has developed into an autonomous medium capable of carrying various types of information. In *Astérix*, for instance, specific typefaces are used to represent different nationalities (e.g. hieroglyphics for the Egyptians, Gothic type for the Goths, etc.). Moreover, the size and width of the letters, the directionality of the lettering (straight, curved, or undulating), letter spacing and vertical orientation as well as the contours of the letters are used to represent the volume, pitch and duration of utterances and noises. Colouring, finally, can be used to symbolise additional information, e.g. about the emotional state of the characters.

These typographic elements are used in comics both in the utterances of the characters and in the representation of noises, through onomatopoeia. In translation, the typographic dimension and its communicative potential have been, and in some cases continue to be, neglected. And yet, the visualisation of acoustic aspects of speech or onomatopoeia may well be used for comic effect, as shown, for example, in a sequence from *Astérix le Gaulois* and its translation into Croatian in Figures 19 and 20.

Asterix teases a Roman by pulling hard on his beard as he sings a nursery song with a line about a beard ("Je te tiens par la barbichette" – "I hold you by your beard"). While the jerky movement of the Roman's head is represented graphically, the resulting trembling of his voice is represented typographically by

Figure 19. Multimodal humour based on graphic and typographic representation in French

Figure 20. Deletion of multimodal humour due to use of standard typography in Croatian

duplicated lettering. The latter was ignored in the Croatian translation, where standard typography suggests a normal voice quality (translation technique C).

6. Summary

Comics must be viewed as complex multimodal texts. All of the verbal and non-verbal elements of this genre can, in principle, be used also to create humorous and comic effects. The repertoire of means of expression in comics includes verbal, pictorial, and typographic signs. In research on humour in general, and in translation studies in particular, the focus has traditionally been on language, with

little or no attention to the impact or functional characteristics of multimodal humour. With regard to comics in particular, translation scholars have investigated only few of the many techniques which can be used in this genre to create humorous and comic effects. Most studies have centred on verbal puns, leaving other factors of verbal humour, such as social stereotyping, irony, reference to other genres, etc. practically unexplored. Humour and comic effects, however, are created not only in the verbal dimension, by establishing new meaning relations outside of communicative norms; para-verbal and non-verbal sign systems also establish comic and humorous relations, by bisociation, violation of communicative expectations and their rich potential for interpretation. Given the growing importance of the visual dimension of texts, translation scholars should take it upon themselves to develop methods of analysis as well as translation procedures for the phenomenon of multimodality, so as to do justice to the holistic nature of semiotically complex texts.

Note

1. Given the large variety of types of comics, it is difficult to come up with a comprehensive definition. What all comics have in common is that they are narrative forms in which the story is told in a series of at least two separate pictures. For a general discussion of the problem of defining comics, see Groensteen (1999). For a translation-relevant anatomy of comics, see Kaindl (1999).

References

Source texts

Binet, Christian (1989). *Monsieur le Ministre*. Paris: Editions Audie. German Translation 1990. *Herr Minister*. Nürnberg: Alpha-comics.

Brétecher, Claire (1979). *Les frustrés 4*. Barcelona: Printer Industria Gráfica. German Translation (1989). *Die Frustrierten 4.* Reinbek bei Hamburg: Rowohlt.

Goscinny, René & Uderzo, Albert (1961). *Une aventure d'Astérix le Gaulois*. Paris: Dargaud. German Translation (1968). *Astérix der Gallier*. Stuttgart: Ehapa. Croatian Translation 1992. *Asterix Gal.* Zagreb: Izvori.

Goscinny, René & Uderzo, Albert (1963). *Astérix et les Goths*. Paris: Dargaud. German Translation I (1965). "Siggi und die Ostgoten." *Lupo modern*, 27–37. German Translation II (1971). *Asterix und die Goten*. Stuttgart: Ehapa.

Goscinny, René & Uderzo, Albert (1967). *Astérix Légionnaire*. Paris: Dargaud. German Translation (1971). *Asterix als Legionär*. Stuttgart: Ehapa.

Goscinny, René & Uderzo, Albert (1968). *Le bouclier d'Arverne*. Paris: Dargaud. German Translation (1972). *Asterix und der Avernerschild.* Stuttgart: Ehapa.

Goscinny, René & Uderzo, Albert (1970). *Astérix en Corse*. Paris: Dargaud. German Translation (1975). *Asterix auf Korsika*. Stuttgart: Ehapa.
Hergé (1947). *Tintin – Le Trésor de Rackham le Rouge.* Paris and Tournai: Castermann. German Translation (1971). *Tim und Struppi – Der Schatz Rakham des Roten*. Hamburg: Carlsen.
Hergé (1958). *Tintin – Coke en Stock.* Paris and Tournai: Castermann. German Translation (1970). *Tim und Struppi – Kohle an Bord*. Hamburg: Carlsen.
Herriman, George (1992). *Krazy Kat*, Vol 2. Wien: Comicforum.
Schulz, Charles M. (1981). "Die Peanuts." *Stern, 47*.

Secondary texts

Attardo, Salvatore (1994). *Linguistic Theories of Humor*. Berlin and New York: Mouton de Gruyter.
Delabastita, Dirk (1993). *There's a Double Tongue. An Investigation into the Translation of Shakespeare's Word-play, with Special Reference to 'Hamlet'*. Amsterdam and Atlanta: Rodopi.
Eco, Umberto (1972). *Einführung in die Semiotik*. München: Fink.
Eco, Umberto (1987). *Semiotik. Entwurf einer Theorie der Zeichen*. München: Fink.
Genette, Gérard (1982). *Palimpsestes: La littérature au second degré.* Paris: Seuil.
Groensteen, Thierry (1999). *Système de la bande dessinée*. Paris: Presses Universitaires de France.
Kaindl, Klaus (1999). "Thump, Whizz, Poom: A framework for the study of comics under translation." *Target, 11* (2), 263–288.
Kloepfer, Rolf (1976). "Komplementarität von Sprache und Bild (Am Beispiel von Comic, Karikatur und Reklame)." *Sprache im technischen Zeitalter, 57*, 42–56.
Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Edward Arnold.
Kotthoff, Helga (1996). *Spass verstehen. Zur Pragmatik von konversationellem Humor*. Wien: Unpublished professorial dissertation.
Lessard, Denys (1991). "Calembours et dessins d'humour." *Semiotica, 85* (1/2), 73–89.
Rabadán, Rosa (1991). *Equivalencia y traducción*. Leon: Universidad de Leon.
Raskin, Victor (1985). *Semantic Mechanisms of Humour*. Dordrecht: D. Reidel.
Wilss, Wolfram (1989). *Anspielungen. Zur Manifestation von Kreativität und Routine in der Sprachverwendung*. Tübingen: Niemeyer.

Chapter 10

Multimodality in operation

Language and picture in a museum*

Andrea Hofinger and Eija Ventola
University of Salzburg, Austria

Museums have traditionally been places where objects are exhibited so that visitors can come and 'view' them. Today, however, museums are spaces where more complex semiotic processes take place. Exhibited objects, visualisations and verbal texts are involved in a dynamic process whereby the visitor interprets his/her experience of the museum. To understand such a process therefore needs multimodal description. This chapter applies multimodal analysis to examples from the Mozart-Wohnhaus (the Mozart Residence) Museum in Salzburg, Austria. The main focus is on the interaction between pictures and spoken language. It will be seen that a closer understanding of how multimodal museum texts work can provide opportunities for museum design, to encourage an integrated interpretative experience for visitors.

1. Introduction

The traditional notion of a museum is a place where artefacts, organised according to prevalent scientific or artistic classifications or associated with historical figures or events, are 'on view'. Visitors may be 'guided' through a museum, either by following and listening to a speaking person, or by reading written texts about the objects' origin, use, connection with a person, event, or historical period, etc. In contemporary museums, however, it is often the case that visitors are immersed in more complex and dynamic processes of experiencing the museum, involving perhaps interactive audio-visualisations, linking spoken or written language with, for example, hearing music, seeing images, moving through display rooms and viewing (or sometimes touching) exhibited objects. The interpretation of these modes and media in interaction forms the visitor's semiotic experience of the museum. An understanding of multimodality in operation is therefore important for making sense of this kind of communication. Texts can support the intended information and assumed perspective(s) of the designers and curators of the museum. The effectiveness of this, however, as well as the extent to which viewers might be

encouraged to engage in their own interpretations, relates to how, and how well, modes and media work together for different audiences.

This chapter discusses multimodal texts from one contemporary museum, in the "Mozart-Wohnhaus" (the Mozart Residence) in Salzburg, Austria. It begins with brief introductions of systemic functional linguistic theory, recent developments in multimodal theory, and previous applications of these to semiotic analysis of museums. The focus then turns to the Mozart Wohnhaus Museum, to considering the global arrangement of rooms, texts, pictures, and objects, before looking more closely at a particular multimodal integration of a Mozart family portrait and an audio-taped recording which explains it and its context.

2. Theoretical background

2.1 Systemic functional theory

The systemic functional linguistic (SFL) model, based on the work of Halliday (e.g. 1978, 1994), is particularly useful in the study of museum semiosis, due to its orientation towards context. SFL approaches texts as (inter)actions in cultural and situational context, by modelling them as cultural types of unfolding social action – *genre* – which have features appropriate to their current social situation – *register*. In other words, texts vary according to who is communicating (*tenor*), about what (*field*) and by what means (*mode*). While these three *register variables* may be realised by both language and other semiotic modes, the study of language has so far been the most developed. SFL theory provides a model of how the situational variables field, tenor, and mode probabilistically activate choices in the linguistic systems, organised by the related metafunctions of language (*ideational*, *interpersonal* and *textual*), and it also provides rich and detailed tools for analysing these choices in text (for a detailed introduction to SFL, see e.g. Halliday 1994; Eggins 1994; Thompson 1996; Martin et al. 1997).

2.2 Multimodality theory

While there are several interpretative frameworks developed for the study of multimodality which provide useful theory and tools, the approaches developed by O'Toole (1994, 1999) and Kress and van Leeuwen (1990, 1996, 2001) will be followed in this chapter, since they share with SFL an orientation towards context and its link-up to semiotic metafunctions, while expanding on the SFL model to account for other kinds of semiosis than language. O'Toole (1994, 1999) applies Halliday's model of the rank scale and the three metafunctions as general semiotic mechanisms for analysing paintings, sculpture, and architecture. He argues

that the artist, like the writer of a text, constructs meanings by choosing options out of the systems of 'Representation' (ideational metafunction), 'Modality' (interpersonal metafunction), and 'Composition' (textual metafunction). The prime concern of this 'functional semiotic model' is to model a visual code which the viewer of an image shares with the artist, and to reveal the functions of particular sign choices in visual art (O'Toole 1994: 215).

Kress and van Leeuwen (1996) also build on Halliday's work to create a descriptive framework for multimodal text. They dismiss Barthes' (1977: 37) idea that images either extend or elaborate the verbal element of a text, or vice versa, stressing rather that

> the visual component of a text is an independently organised and structured message – connected with the verbal text, but in no way dependent on it: and similarly the other way around. (Kress & van Leeuwen 1996: 17)

Kress and van Leeuwen, as well as O'Toole, also regard an image not only in aesthetic terms but look at the dynamic interplay between a painting (or other text) and the social situations in which it is created and used. (Kress & van Leeuwen 1996: 17; O'Toole 1994: 216). Multimodality, as defined by Kress and van Leeuwen (2001: 20), is the use of several semiotic modes and their combination within a socio-cultural domain which results in a semiotic product or event. This definition works admirably for the design and interpretation of museums.

2.3 Linguistic and multimodal analyses of museums

Useful work on linguistic and multimodal analysis in museum contexts has been done by Ravelli (1996, 1997, 1998) and Purser (2000). Ravelli (1997: 6) explores the basics of meaning-making in museums. She stresses that meaning is not to be equated with content, but, rather, adopts Halliday's metafunctions to look at all museum semiosis as built from ideational ('content') meanings in conjunction with interpersonal and textual ones. She sees choices from systems such as *relationship*, *contrast*, *difference*, and *similarity* as essential to meaning-making; that is, an object is meaningful only in relation to other (sets of) objects (Ravelli 1997: 3). At some level, everything in a museum carries meaning, through both **semiotic systems**, e.g. of language or images, and *intersemiotic systems*, e.g. relating an image to a verbal text (Ravelli 1997: 5). As Ravelli (1997: 10) puts it, the fundamental point is "that *all* meaning is constructed, and that all meaning has a specific socio-cultural location, whether one is aware of it or not." By selecting objects and making choices about displaying them, e.g. according to particular orderings or groupings, a museum attaches cultural significance to them; but meaning is also negotiable, and may depend on the previous experiences of different interactants (Ravelli 1997: 6).

While we are sometimes unaware of the meanings made around us, Ravelli (1997: 9) also points out that we tend to first notice those meanings with which we do not agree. This is the case in the museum example discussed by Purser (2000), based on an informant's response to the representation of indigenous people, in an exhibition in Berlin about the South Pacific. While, in her view, the museum portrays these people in biased way, a viewer who is not 'in the know' about the subject matter is likely to accept the museum's meanings as "natural, obvious and true" (Purser 2000: 185). Purser is therefore interested in the role of museums in the public ideological and educational arena; for example, in the representation of one culture by another, and what kind of 'voice' the represented people (of the exhibition) have to speak for themselves in their representation by a museum. Rather than taking an intersemiotic approach, like Ravelli (1997), Purser investigates the role of language as a support for visuals in the museum. Objects in a museum need some interpretation, and verbal museum texts represent "facts the museum claims to know", and are also means "to frame events and people in particular ways" (Purser 2000: 175). Like Ravelli, Purser (2000: 175) argues that choice and selection are essential for the meaning-making of a verbal museum text, which she calls a "record of selection" from a potential of meanings. These choices imply contextual motivations, consequences and ideology: "Even as we speak 'about' someone else, we are representing ourselves, our values, our desires, the ways we want to tell our story" (Purser 2000: 185). This approach to museum semiosis adopted by Ravelli and Purser, and the social semiotic model of language and multimodality outlined in the preceding section inform our discussion of the Mozart-Wohnhaus Museum, in Section 3.

3. The questions to be explored

Our investigation of examples from the Mozart-Wohnhaus Museum will consider choices in language and image and how they interrelate. In our analysis and discussion, we are concerned with the following questions: What meanings are made – what content, what engagement with the visitor, what organisation or prioritisation? Which semiotic (linguistic/visual) choices have been made to express these particular meanings? What do these tell us about the ideology and cultural context of the museum? To pursue these questions, Section 4 first introduces the museum context to the readers. Section 5 then presents the text and the picture chosen as data and the results of the linguistic and visual analyses. This is followed in Section 6 by a synthesis of the analyses and discussion, and Section 7 concludes the chapter.

4. An overview of Salzburg's Mozart-Wohnhaus Museum

Salzburg is known all over the world as the birthplace of Wolfgang Amadeus Mozart, and it is no surprise that the city both praises its famous son and cashes in on his fame. As well as the Mozart-Wohnhaus Museum which is examined in this chapter, Salzburg has a second Mozart museum, the Mozart-Geburtshaus Museum, where the composer was born. In the Mozart-Wohnhaus Museum, emphasis is given to Mozart's life and work during the years 1773-1780. Rather than presenting the common image of the musician as a child prodigy and musical genius, the museum presents Mozart as a son, brother, lover, and friend, through a set of relationships with friends and family. As well as through displays of objects and pictures of Mozart, his family and friends, this representation is realised through recorded spoken texts. Every visitor is offered a personal audio-player for these recordings, which are activated in particular zones of the museum by infra-red sensors. That is, when the visitor comes to a particular display area, a spoken text relevant to that section automatically plays. (The audio-recording also includes examples of Mozart's music. While we recognise the importance of the music to the multimodal experience of the museum, an analysis of these musical extracts will not be attempted here.)

4.1 The floor plan of the museum

The museum consists of seven rooms. Figure 1 displays the order of these rooms. The first large room in the left-hand corner is called "The Dancing Master's Hall" which was used during Mozart's time for parties, sitting together and playing games. The second room introduces some of Wolfgang's employers and supporters, important women in his life, and his connection to the church. The third room, known as "The Library of Leopold Mozart", illustrates the personality of Wolfgang's father. The fourth room, in the right-hand corner, is dedicated to Maria Anna Mozart, or 'Nannerl', Wolfgang's sister. The fifth room aims to convey an atmosphere of Mozart's family life and contains furniture which dates back to Mozart's time. In the sixth room the visitors can trace the musician's numerous travels on a wall-map. Finally, the last room offers a slide show which gives an outline of Mozart's career and once again refers to his relationships to family and friends. The following analysis will focus on The Dancing Master's Hall, the centre of the Mozart family's social life.

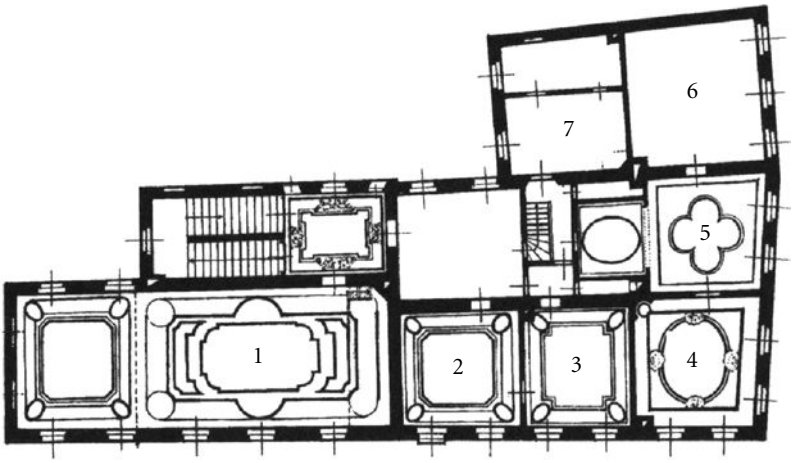


Figure 1. Floor plan of the Mozart-Wohnhaus Museum: (1) Dancing Master's Hall, (2) Employers and supporters, (3) Library of Leopold Mozart, (4) Room of Nannerl Mozart, (5) Furniture, (6) Travels, (7) Slide show

4.2 Zones in The Dancing Master's Hall

The Dancing Master's Hall is a big, festive room, which was used as a kind of a family living room and as a room for receiving guests. On one side of the room there are four glass display cases containing silhouettes, notes written by Mozart, and small paintings. On the other side of this room, there are several of the musical instruments which Wolfgang Amadeus Mozart played (see Figure 2 and Figure 3). Just above the instruments one sees a very large Mozart family portrait, (the focus of the visual analyses in Section 5.2).

If the visitor has chosen to walk through the museum by listening to the audio-recording, s/he will find that there is a spoken text for six different infra-red zones in The Dancing Master's Hall, corresponding to the space in front of individual display cases, musical instruments or pictures, each telling a story or giving details about part of Mozart's life. Zone 1 introduces the visitor to the room and deals with a particular game played in Mozart's time. Zone 2 covers musical instruments and the large family portrait which hangs above them. Zones 3 and 4 describe antique instruments on display. Zone 5 concentrates on Mozart's compositions, and Zone 6 is about friends of the Mozart family. Of these, Zone 2 (the instruments and the family portrait) is the focus of our linguistic and multimodal analyses in the next section, and we look particularly at how the audio-text and the picture work together.

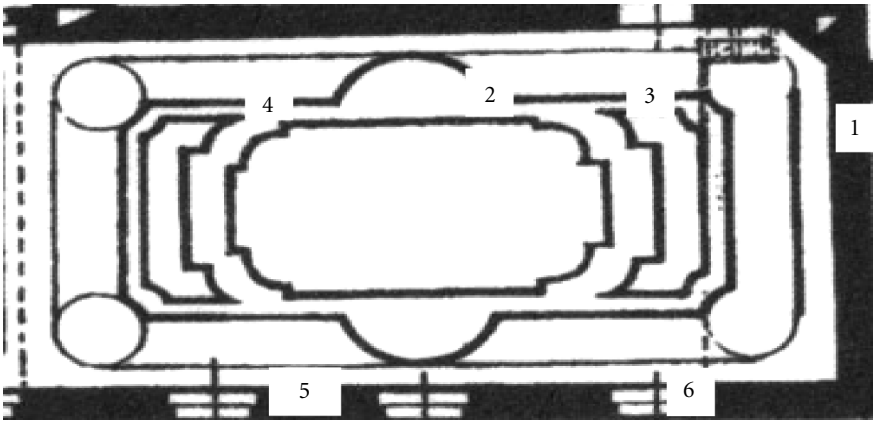


Figure 2. Zones in The Dancing Master's Hall: (1) Introduction to the room, (2) The family portrait, (3) Instruments, (4) Instruments, (5) Compositions and dedications, (6) The 'Haffner-family'

Figure 3. The Dancing Master's Hall, Festschrift, Picture 22, p. 75 (© Internationale Stiftung Mozarteum (ISM))

5. Analyses

5.1 Linguistic analysis of the text

As mentioned above, family, friends, and the surroundings of Wolfgang Amadeus Mozart play a major role in the ideational meanings made in the Mozart-Wohnhaus. A transcription of the audio-text for Zone 2 is given below as Text 1.

(1) In the late autumn of the year 1780, shortly before Wolfgang left to go to Munich for the premiere of his opera Idomeneo, a painter, still unknown to us nowadays, started to paint the only authentic portrait of the family. The work was probably commissioned by Leopold Mozart. We know from a series of letters that Wolfgang sat for the painter shortly before he went to Munich; Nannerl sat at the end of the year and Leopold himself was probably not painted until his return from Munich. The family portrait shows father Leopold holding his violin, and Nannerl and Wolfgang playing the piano. The portrait of their mother, who had died two years earlier in Paris, is hanging on the wall. The family portrait was for a long time ascribed to Johann Nepomuk Della Croce. Mozart's sister Nannerl wrote some years later about the painting that it showed the best likeness of the brother. There are two key instruments in front of the picture. The one on the right is an Italian harpsichord from the 17th century on which you will now hear a piece from the Balletto by Bernardo Storace.

5.1.1 *The ideational metafunction*

The ideational metafunction of language is that of representation of 'things in the world' – participants, processes, relationships, etc. Analysis of the lexicogrammatical system of transitivity sheds light on this metafunction in the museum texts. As Purser (2000: 177–178) puts it, "transitivity refers to the language choices made to represent 'who is doing what to whom' in the image of the world a given text creates".

The analysis of transitivity in Text 1 (see Appendix 1) shows that most of the processes are material (ten in nineteen clauses). There are six relational processes, two mental processes, one verbal, and one existential process.

The material processes convey a feeling of activity and movement in the text. They tell the visitors about the production of the family portrait: who commissioned the painting, what the painter did, when the Mozarts sat for the portrait, and what actions are represented in the painting. There are five actors in the text: Wolfgang, Nannerl (the sister), Leopold (the father), Anna (the mother), and the painter. There is no specific predominance of Wolfgang Amadeus Mozart's actions, representing him merely as another member of the family. The material processes

involving Mozart are 'going to Munich', 'sitting for the painter', and 'playing the piano with his sister'. Nannerl's actions involve 'playing the piano' and 'sitting for the portrait'. The father, Leopold Mozart, acts in 'commissioning the portrait' and in 'holding the violin', while the mother's only action is 'dying'. Finally, the painter's material actions are 'painting the portrait' and 'painting Leopold'.

Speaking of relational processes in museum texts, Purser (2000: 181) says that they "are used to locate, define, classify and evaluate" objects, people, and activities. In Text 1, the six relational processes define the portrait and what it represents: 'the portrait shows the family', 'the portrait of the mother is hanging on the wall'. Another relational process offers the visitor an interpretation of the visual reality of the picture: 'the family portrait shows the best likeness of Mozart'. The sources of information for the textual and visual reality are represented through one of the two mental processes: the museum curators 'know from a series of letters' facts about the production of the portrait. Nannerl is also made an information source, confirming the likeness between Mozart and his representation in the portrait, as the Sayer in a verbal process: "Nannerl wrote [. . .] that it showed the best likeness of her brother". An additional relational process identifies the painter of the picture as unknown to us.

Towards the end of the audiotape text, an interesting multimodal shift is achieved. The existential clause, "There are two key instruments in front of the picture", draws the visitor's focus away from the picture to the instruments. It then identifies one of the objects in a relational clause, "The one on the right is an Italian harpsichord", but leaves the other unidentified. The multimodal experience then rushes to a change from mental 'viewing' to 'hearing': "you will now hear a piece from the Balletto by Bernardo Storace". The visitor's experience moves from listening to the audio-text to viewing the portrait and then the harpsichord and back to listening to a piece of music.

In summary, the audio-text which accompanies the portrait mainly lists the major figures represented, leaving the visitor to engage with the other meanings of the painting on her/his own. The visitor's focus is then moved out of the picture in a rather abrupt way by introducing the two instruments in front of the picture. The last processes link the physical harpsichord in the museum with another modal realisation, music.

5.1.2 *The interpersonal metafunction*

As the museum offers information rather than demands it, the interaction between the museum and the visitor is reduced to 'non-negotiation' rather than 'negotation' of meanings (or 'monologuing' rather than 'dialoguing'), and the mood choices in the audio text are restricted to declaratives. The Mozart-Wohnhaus does not make use of imperatives, although this is sometimes the case in other muse-

ums; e.g. suggesting what the visitors should do mentally: "Look at. . ." (cf. Enkvist 1991). However, there are other kinds of interaction in the Mozart-Wohnhaus museum between the museum as an addresser and the visitor as audience. Looking at the audio text, we find that some uncertainty is expressed in the information given to the visitor. The recorded speaker uses the mood adjunct *probably* twice, to express a median value of probability. In this way, the museum withdraws some of its responsibility for the truth value of the message: "The work was probably commissioned by Leopold Mozart"; "Leopold himself was probably not painted until his return from Munich".

5.1.3 *The textual metafunction*

The textual metafunction relates to how a text is organised and structured. Theme, the first element in the clauses, provides the starting points for visitors, and often the given, or known, information. The later part of each clause, the Rheme, captures what is 'news' for the listener. Text 1 starts out with a marked theme realised by a circumstantial adjunct of time, *in the late autumn of the year 1780*, the period and circumstances when the portrait was painted. In fact, time plays an important role in structuring the text. Some examples are: *at the end of the year*, *two years earlier*, *some years later*. However, these expressions of time are not used in the most efficient way for structuring the text – i.e. thematically; only *shortly before* appears in a thematic position. More often (in three clauses) the painting itself, the 'family portrait' or 'the work', realises an unmarked theme, which reveals the intention of making the visual in front of the visitor a starting point for structuring the verbal text.

An existential theme *there* towards the end of the text, *There are two key instruments*, breaks the structure of the text around the family portrait and introduces a new topic by drawing the visitor's view to a specific place outside the portrait. This is further located by the last theme: *The one on the right*. In this way, the museum again stresses the visual element. This new introduction to musical instruments might also be seen as a link to the following two zones, which deal with music in more detail.

The brief Text 1 of Zone 2 offers basic information about the family portrait and gives the impression that the museum itself has only little background information about it. At this point, we can raise the question: should the museum's text focus more on meaning in the painting itself – its structures, colours, and techniques – thereby opening the opportunity for visitors to engage more directly with the visual text, in line with the position taken by O'Toole (1994)? Visuals provide a large part of museum meanings, as Section 5.2 will exemplify. What role can a museum play in guiding visitors' interpretations of visuals – or is it up to the individual viewer to engage with the picture on his/her own?

Figure 4. Mozart family portrait in The Dancing Master's Hall (Festschrift, Picture 4, p. 59) (© Internationale Stiftung Mozarteum (ISM))

5.2 Visual analysis of the family portrait

The family life of the Mozarts plays an important role in the Mozart-Wohnhaus museum, so the family portrait in The Dancing Master's Hall may be regarded as an important artefact. Indeed, the picture of Wolfgang with his family (Figure 4) embodies the central concept of this particular museum – Mozart as a family member and as a friend – while also showing him at the piano in line with his more familiar musical identity. The following analysis of this painting will be according to the three semiotic metafunctions, following O'Toole (1994, 1999) and Kress and van Leeuwen (1990, 1996, 2001).

5.2.1 *The interpersonal metafunction*

The interpersonal metafunction, (or the engagement function, in O'Toole's terms), is reflected in the way pictures capture our interest and involvement with the depicted subject(s). O'Toole (1999) suggests viewers should ask themselves what engages them most when looking at an image. In the Mozart family portrait, the most striking elements are the direct gaze of the participants, the bright red of Mozart's jacket, the grouping of the family members and Wolfgang's po-

sition in the middle with the piano. Another eye-catcher is Mozart's right hand, which is placed over Nannerl's left hand. This particular feature links the two siblings, but also shows Wolfgang's relative dominance in the painting. His hands also function as vectors to the piano which is situated right in the middle of the painting, emphasising the importance of its link with him; a link which the audio text strengthens by drawing the visitor's attention to the musical instruments in front of the picture.

All four represented participants engage the viewer and create affinity by gazing directly outward and smiling (Kress & van Leeuwen's 1996: 129). According to Kress and van Leeuwen (1996: 143–145), the horizontal axis of a painting reflects involvement. The museum visitor looks at the family portrait from a frontal viewpoint which gives her/him the feeling of being involved in a part of Mozart's world. The vertical angle, in contrast, usually expresses power (Kress & van Leeuwen 1996: 146–147). As the family is seen neither from above nor below, there is no power difference construed between the visitor and the participants of the painting; they are on the same level. The size of the frame indicates social distance between the viewer and the represented participants (Kress & van Leeuwen 1996: 130). The family portrait is a 'medium shot'; that is, the participants are not portrayed full size, and this seems to bring them socially closer, making the social distance fairly informal, but not intimate.

In sum, whereas the audio text accompanying the family portrait does not encourage much interaction, the interpersonal analysis of the visual portrait shows that the painter makes choices which do create engagement with the viewer.

5.2.2 *The ideational metafunction*

The ideational (or representational) metafunction of a painting is revealed in the people, objects, events, and landscapes depicted. In Kress and van Leeuwen's terms (1996: 63), the family portrait is a narrative, transactional process. The actions are those described in the audio text: "Wolfgang Amadeus Mozart is playing the piano together with his sister Nannerl. His father is sitting next to him, holding the violin". In contrast to the audio text, in the painting Wolfgang is, however, obviously the main actor, as indicated by his position in the middle of the episode, the high colour saturation of his red jacket, and the contrast between the jacket and the dark browns of the background. He additionally has 'psychological salience' (Kress & van Leeuwen 1996: 61); the visitors of the museum automatically look at Wolfgang as it is he who is the reason for their visit. Another contrast between the image and the spoken text is that the portrait represents persons and action in the circumstance of a room, i.e. in space, while the audio text actions take place in circumstances of time. The space of the painting is echoed in the physical museum

space of the Dancing Master's Hall, which was used for the same actions of sitting together and playing.

5.2.3 *The textual metafunction*

In the painting's textual (or compositional) structure, a certain unity is expressed by a triangular frame which starts on top of the mother's portrait and encloses the rest of the family. Thus, the portrait of Anna Mozart, itself framed, constitutes the starting point for the frame of the family portrait. Placed right in the middle, it divides the family into two halves; Nannerl and Wolfgang sit as a unit on the left side, with Leopold on the right. However, the three participants are also connected by other devices. They are more or less at the same height, and a strong horizontal line runs through the picture from Nannerl's head-band on the left, over the top of Wolfgang's head, touching the scarf tied at Anna's throat in the small portrait, over Leopold's brow and ending at the tip of the violin on the right. There is a parallel line below, from Nannerl's elbow to the piano. A few smaller parallels are established by the violin bow, the piano horizontals, and the bottom of the music stand. As a balance to these horizontals, verticals are formed by the erect bodies of the participants, the beams on the wall and the edges of the piano. Diagonals from the lower left to the upper right bring some movement into the painting, through the music stand, the violin and the frills on Nannerl's dress. There are complementary diagonals from the upper left to the lower right, the upper arms of Nannerl and Wolfgang, the back of the armchair, and the frills on Nannerl's neckline.

This analysis shows that the basic compositional principle of the painting is balance and harmony. Although the viewer first concentrates on Wolfgang in the middle, the overall effect is to present the Mozart family as a harmonious whole. This compositional structure therefore aligns perfectly with the museum's overall concept.

6. Synthesis and discussion

To summarise, the multimodal analyses of Zone 2 of the Dancing Master's Hall in the Mozart Wohnhaus Museum reveal that language and visuals construe both similar and different meanings. The audio-text, Text 1, offers some basic information on the period, the production, and the representation of the family portrait, and partially guides the visitor's focus between audio and visual information. The visual analyses of the portrait reveal a number of available meanings, some of which support the audio-text, while others are complementary. However, the museum gives very little guidance to the viewing and interpretation of this painting.

More could be done to help the viewer in his/her processes of experiencing this part of the museum. In our opinion, interaction between the museum as an addresser and the visitor as an addressee could be improved by re-structuring the information in the audio-text. Looking at the audio-text in terms of stages which guide the visitor's semiotic experience shows specific opportunities for improvement. Below is a re-written version of Text 1. We propose that this Text 2, with its clear ideationally, interpersonally, and textually organised structure, might possibly better integrate language and visuals for visitors to the museum.

(2) **I. Orientation (mental)**

1. Now please look at the Mozart family portrait in front of you.

II. The picture and the painter (relational)

2. This painting is the only authentic portrait of the family.
3. The painter is still unknown to us today,
4. although for a long time it was ascribed to Johann Nepomuk Della Croce.

III. The picture and its production (material)

5. The family portrait was probably commissioned by Leopold Mozart.
6. We know from a series of letters when it was painted.
7. Wolfgang sat for the painter in the late autumn of the year 1780,
8. shortly before he went to Munich for the premiere of his opera Idomeneo.
9. Nannerl sat at the end of the year
10. and Leopold was probably not painted until his return from Munich.

IV. The picture and its representation (relational and material)

11. The family portrait shows
12. Father Leopold holding his violin,
13. and Nannerl and Wolfgang playing the piano.
14. The mother,
15. who had died two years earlier in Paris,
16. is represented by her picture on the wall in the family portrait.
17. Some years later Mozart's sister Nannerl wrote about the painting
18. that it showed the best likeness of her brother.

V. Moving away from picture to the other objects (relational and existential)

19. Let us now focus on what is in front of the picture.
20. There are two key instruments.
21. The one on the left is . . .
22. and the one on the right is an Italian harpsichord from the 17th century

VI. Focus on music (mental)

23. You will now hear a piece played on it from the Balletto by Bernardo Storace.

The original museum text neither guides the visitor into looking at the picture nor looking out of it. In Phases I, V, and VI, all involving mental processes, the new text, Text 2, aims to build in guidance for the visitor's semiotic experience, from the picture to the physical room to the music. We propose that the origin of the object and its maker could then provide a starting point for considering the painting. Therefore, in Phase II, information on the painting and the painter is brought together (largely in relational clauses), while in the original it was spread here and there. Phase III then covers the material processes of the painting's production: who sat as a model for the painter, and when. The representational aspects of the portrait (relational and material processes) follow in Phase IV. (These could perhaps be extended, to include suggestions for interpreting the visuals.) In Phase V, the rewritten version gives an explicit cue for the shift in visual focus (existential and relational processes), while Phase VI achieves the final transition from the visual to music.

7. Conclusion

This chapter has looked at multimodal meaning-making in the Mozart-Wohnhaus Museum in Salzburg, and how such a process is aided by the audio-text offered to visitors. Understanding multimodality in operation allows new opportunities for museum designers and curators. The museum will offer the viewer certain perspectives through its written or spoken texts, displays, and other semiosis. Visitors will understand, accept, or reject the proffered views on the basis of their own interpretations of the whole semiotic experience of the museum. Museums can take a more active part in encouraging this dynamic process through awareness of how different kinds of semiosis work together.

Note

* The authors thank the Internationale Stiftung Mozarteum for their permission to reproduce the figures.

References

Angermüller, Rudolph, Schlie, Reimar, & Sertl, Otto (1996). *Festschrift. Die Wiedererrichtung des Mozart-Wohnhauses. 26. Jänner 1996.* Salzburg: Internationale Stiftung Mozarteum.

Barthes, Roland (1997). *Image, Music, Text.* London: Fontana.

Eggins, Suzanne (1994). *An Introduction to Systemic Functional Linguistics.* London: Frances Pinter.

Enkvist, Nils Erik (1991). "Discourse strategies and discourse types." In E. Ventola (Ed.), *Functional and Systemic Linguistics. Approaches and Uses* (pp. 3–22). Berlin and New York: Mouton de Gruyter

Halliday, M. A. K. (1978). *Language as Social Semiotic.* London: Arnold.

Halliday, M. A. K. (1994). *Introduction to Functional Grammar.* London: Edward Arnold.

Kress, Gunther & van Leeuwen, Theo (1990). *Reading Images.* Victoria: Deakin University Press.

Kress, Gunther & van Leeuwen, Theo (1996). *Reading Images. The Grammar of Visual Design.* London and New York: Routledge.

Kress, Gunther & van Leeuwen, Theo (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication.* London: Arnold.

Martin, James R., Matthiessen, Christian M. I. M., & Painter, Clare (1997). *Working with Functional Grammar.* London: Arnold.

O'Toole, Michael (1994). *The Language of Displayed Art.* London: Leicester University Press.

O'Toole, Michael (1999). *Engaging with Art. A New Way of Looking at Paintings.* Perth: Murdoch University.

Purser, Emily (2000). "Telling Stories: text analysis in a museum." In E. Ventola (Ed.), *Discourse and Community: Doing Functional Linguistics* (pp. 169–198). Tübingen: Gunter Narr Verlag.

Ravelli, Louise (1996). "Making language accessible: successful text writing for museum visitors." *Linguistics and Education, 8* (4), 367–387.

Ravelli, Louise (1997). "Making meaning: how, what and why?" In *Museums Making Meaning: communication by design*. Papers from the Museums Australia Inc (NSW) 1997 State Conference (pp. 1–11). Sydney: Museums Australia.

Ravelli, Louise (1998). "The consequences of choice. Discursive positioning in an art institution." In A. Sánchez-Macarro & R. Carter (Eds.), *Linguistic Choice Across Genres* (pp. 137–153). Amsterdam and Philadelphia: Benjamins.

Thompson, Geoff (1996). *Introducing Functional Grammar.* New York: Saint Martin's Press; London: Hodder Education.

Floorplans and images

Angermüller, Rudolph, Schlie, Reimar, & Sertl, Otto (1996). *Festschrift. Die Wiedererrichtung des Mozart-Wohnhauses. 26. Jänner 1996.* Salzburg: Internationale Stiftung Mozarteum.

Appendix 1

Zone 2 – Processes in the original Residence text

1. "In late autumn of the year 1780,
2. shortly before Wolfgang left to go [material] to Munich for the premiere of his opera *Idomeneo*,
1. a painter,
3. «(who is [relational]) still unknown to us nowadays»,
1. started to paint [material] the only authentic portrait of the family.
4. The work was probably commissioned [material] by Leopold Mozart.
5. We know [mental] from a series of letters
6. that Wolfgang sat [material] for the painter
7. shortly before he went [material] to Munich;
8. Nannerl sat [material] at the end of the year
9. and Leopold himself was probably not painted [material] until his return from Munich."
10. "The family portrait shows [relational]
11. Father Leopold holding [material] his violin
12. and Nannerl and Wolfgang playing [material] the piano.
13. The portrait of their mother,
14. «who had died [material] two years earlier in Paris»,
13. is hanging [relational] on the wall."
14. The family portrait was for a long time ascribed [relational] to Johann Nepomuk Della Croce.
15. Mozart's sister Nannerl wrote [verbal] some years later about the painting
16. that it showed [relational] the best likeness of her brother.
17. "There are [existential] two key instruments in front of the picture.
18. The one on the right is [relational] an Italian harpsichord from the 17th century
19. on which you will now hear [mental] a piece from the *Balletto* by Bernardo Storace [pronounced: Storatsché]."

Chapter 11

Drawing on theories of inter-semiotic layering to analyse multimodality in medical self-counselling texts and hypertexts

Eva Martha Eckkrammer
University of Salzburg, Austria

Ever since different modes and media started to enlarge our textual cosmos by combining and intertwining, multimodality has become a crucial issue in linguistic description. However, there are few attempts to methodically categorise multimodal "blending" in text, or in hypertext. This chapter, which is grounded in the framework of contrastive textology, focuses on the value and role of theories of inter-semiotic layering, particularly in regard to comparative multilingual text and corpus analysis. I begin by addressing some preliminary issues for the intersection of multimodality and corpus linguistic research. This is followed by a discussion of the concept of hypertext, in contrast to text, with its multiple possibilities for blending modalities, and the implications for theories of multimodality. The model of inter-semiotic relationships and layering proposed by Hoek (1995) provides a useful approach, which could fruitfully be integrated with other contemporary perspectives (such as Kress & van Leeuwen 2001). This model is applied here to samples from the DIALAYMED-corpus, a multilingual genre corpus consisting of medical self-counselling texts on infectious diseases, to illustrate the different forms of multimodality found in current (hyper)texts. Applying the theoretical approach suggested in this chapter to practical analysis highlights several issues which need to be elaborated in future multimodality research, particularly with regard to text-intelligibility.

1. Introduction

The 20th century was characterised by a formidable increase in the number of ways information can be stored, conveyed and retrieved – from moving pictures to telecommunication and broadcast media and finally computers and their world-wide connection, the internet. These channels which currently carry a large proportion of human communication continue to develop and expand, both substituting and complementing traditional forms. Particularly in hypertextual environments, but also in traditional texts, semiotic modes are combined, blended and

intertwined to a growing extent. The question arises whether it is possible to grasp these contemporary textual phenomena with theoretical devices long established in linguistics, or whether we need to develop an expanded model of inter-semiotic layering. As Kress and van Leeuwen (2001) have shown, it is not enough to look at the semiotic resources of texts alone, if we aim at a multimodal approach to communicative products. That is, if our point of departure for textual analysis is a concept of genre as socially functional practice, then all aspects of the discourse community, the media, the production, design and distribution of the text, and finally its reception and interpretation, need to be taken into account.

Accordingly, this chapter takes as a starting point experiences in comparative approaches to genre-based multilingual text corpora and suggests a pragmatic approach to inter-semiotic layering that might be applicable to all forms of multimodal texts in traditional print or hypertextual environments, and more particularly relevant to corpus linguistics. After some theoretical consideration of how to pave the way for a deep analysis of multimodal text corpora in Section 2, Section 3 briefly addresses the question of whether we need a new conception of text when including hypertextual samples in the analysis. Section 4 then focuses on the different forms of inter-semiotic layering modelled by Hoek (1995), and tests this model for adaptation to hypermedia environments by applying it to text samples from the DIALAYMED-corpus (a diachronic corpus of lay-person oriented medical texts on major infectious diseases, currently compiled in Salzburg). Selected examples are provided to illustrate implications for future challenges and opportunities in multimodal research, for instance with regard to intelligibility theory.

2. Paving the way for a deep-analysis of multimodality in contemporary text corpora

Traditionally, corpus linguistics has paid little attention to non-verbal elements in texts. The reasons are obvious in light of the shift towards corpus methods which foreground the development and application of automated analysis. In this respect, I perfectly agree with Kennedy (1998) who stresses that corpus linguistics is basically about asking and answering insightful linguistic questions and not an end in itself:

> Although there have been spectacular advances in the development and use of electronic corpora, the essential nature of text-based linguistic studies has not necessarily changed as much as is sometimes suggested. [...] Corpus linguistics did not begin with the development of computers but there is no doubt that computers have given corpus linguistics a huge boost by reducing much of the drudgery of text-based linguistic description and vastly increasing the size of the databases

> used for analysis. [...] It should be made clear, however, that corpus linguistics is not a mindless process of automatic language description. Linguists use corpora to answer questions and solve problems. Some of the most revealing insights on language and language use have come from a blend of manual and computer analysis. (Kennedy 1998: 2f.)

This point seems particularly important when extending corpus linguistics to embrace multimodality in text and hypertext. If we refer to current tools that allow electronic analysis (see for example the software reviews by Alexa & Zuell 1999) the tools for integrative multimodal analysis are still very limited and extremely time-consuming. However, the development of standards such as SGML and XML seem to pave the way to handling text as a semiotic whole, including non-verbal elements (cf. Bateman, Delin, & Henschel in this volume). In any case, we have to be aware that an integrative approach to text, considering the verbal and visual as equally important, and perhaps even including a time-space-axis, entails considerable change in the size of the corpora needed. It therefore seems crucial at this point to briefly address the question of why it is important to dedicate time and resources to multimodal aspects of (hyper)text.

There are numerous ways in which language, verbal text in written or spoken form, may be related to static and/or moving pictures. The pictorial has obviously played an essential role both before and after the advent of literacy. Indeed, there are several cultures that still give prominence to the oral and pictorial modes and do not place the same emphasis on written modes as has been established in the modern Western societies. More to the point, the pictorial and verbal modes have never excluded each other but have always intertwined to form a semiotic texture in text. This fact, of course, does not spare us from questions about the precise nature of this semiotic intertwining. In what ways may a linguistic item be contiguous to, or form part of, a pictorial representation, or vice versa, to give rise to particular forms of semiotic interaction?

I will call these forms of semiotic interaction *inter-semiotic layering* or *imbrication*,[1] as they may (but do not necessarily) give prominence to either the verbal or pictorial, and may also include a time-space axis in hypermedia environments. It is clear that inter-semiotic layering can include all the types of signs distinguished by Peirce and his followers (icons, indices and symbols). Within inter-semiotic layering the functional spectrum of pictorial information, which was for a long time rather restricted to illustration and explanation, is nowadays wider and includes, among others, all Jakobsonian functions (referential, emotive, conative, metalinguistic, phatic or narrative-aesthetic). Furthermore, the blending of the verbal and pictorial can incorporate playful elements and evoke superimposed meanings to its referent, when it exercises a symbolic function. In fact, the non-verbal elements, particularly in hypermedia, play a significant role in the analysis of functional hier-

archies in texts. They can detail certain semantic aspects, reveal feelings and values transported by the addresser, add humour, try to influence the behaviour of the addressee, etc. In certain cases (moving) pictures can function similarly to punctuation marks in writing, i.e. to indicate the beginning or the end of a paragraph or text, create a pause in the reception, or claim additional attention (for instance in advertisements). It should therefore be emphasised that pictorial and verbal elements in texts are bound together and never develop their meaning separately, but do so through a process of interaction between the semiotic layers. Images tend to perform multiple functions strictly bound to their verbal surroundings, and vice versa. However, functions organise themselves hierarchically in relation to a dominant function which is usually bound to the genre. A hierarchical model of functions is, at first, built on the two highest functions of language and text, viz. the communicative and cognitive function. Subordinate to these main functions we find a series of functions such as those related to the different elements of the language system, i.e. the functions of syntax and morphology as discoursed-driven elements of language which lead to complex clause structures and words and serve textual cohesion (cf. Dressler & Eckkrammer 2001).

Interestingly enough, the issue of verbal-visual interaction has traditionally been formulated in terms of rules and principles, based on an assumption that our interpretation of pictures is mediated by our linguistic competence. If we reflect on the historical and cognitive dimensions of this approach, however, the opposite conception seems far more plausible – the production and reception of linguistic elements appears to be strongly influenced by images. Iconicity, even if paradoxically refuted by prominent semioticians such as Eco, seems to be a significant factor with regard to the naturalness of language (cf. Dressler 1999). The approach taken here follows a text model of natural text linguistics (cf. Dressler 1989, 2000) that includes the notion of preferences (cf. Dressler 1999), set by cognitively-based semiotic parameters, such as iconicity, indexicality, transparency and contrast between figure and ground. These parameters affect both the text producer and the interpreter, i.e. when choosing a way to combine the verbal and the pictorial during text production or making meaning of a multimodal text. The described text model draws on a cognitive, pragmatic, functional and communicative conception of text that, as the next section will illustrate, is useful in contrastive analysis which encompasses both text and hypertext.

3. Text versus hypertext: Do we need a new conception of text?

When looking at multimodal discourse from a comparative perspective, dealing with a corpus that includes not only various languages, but also (hyper)text across

different media, traditional conceptions of text are hard-pressed to account for the full range of multi-semiotic and highly networked textual phenomena. To meet this situation, hypertext may be usefully defined as

> a concept and new means for structuring and accessing text in distance-communication, based on software-technology which allows and suggests the interconnection of text by means of electronic links. The elements can be independent documents (nodes) or different sequences of one and the same document. The links can be either internal (directed to nodes within the same hypertext) or external (pointing to other hypertexts).
>
> (Adapted from Storrer 2000; Engebretsen 2000)

Hypertext is not a genre, but may realise many functionally differing genres embedded in numerous social practices, or discourses. Hypertext favours non-linear structures in text and, through its computational medium, enables the use of more than one mode and overlapping semiotic codes. Multimodality is fostered by hypertextual environments just as much as the chunking and interconnection of texts through visible electronic relations (links). Therefore, the 'reading' of hypertext is not necessarily bound to a particular receptive chain, but permits the reader to choose and create an individual path to construct a continuity of meaning. Additionally, hypertexts may dissolve the clear-cut line between the text producer and the reader by fostering interactive processes. The interconnectedness and non-linearity of hypertextual constructs imply that the "comprehension and discursive structure (. . .) is volatile to the extent that it is pragmatically, not grammatically, determined, and so remains outside of the normative prediction and pattern" (Miles 2000). The reception process and its result can therefore hardly be predicted.

It is beyond the scope of this chapter to develop a fully elaborated typology of hypertext and hypertextual linkage, although various criteria have been proposed on which to base such a classification. An important example is sequentiality (see for instance Storrer 2000, who distinguishes between mono-, poly- and unsequential hypertext). There is also general agreement that the classification of hypertextual linkage requires an integrative approach which takes a number of variables into account, such as function, position, visual elements, etc. While there is a widely accepted need to expand hypertextual analysis to take into account both different semiotic levels and cognitive processes, the underlying conception of text – so long as it is a communicative and pragmatic one (such as that following de Beaugrande & Dressler 1981) – still seems to provide the foundation. The constitutive criteria of textuality, such as coherence, cohesion, situationality, acceptability, intentionality, informativity and intertextuality may be developed, but appear sufficiently dynamic to adapt to hypertext. A concept of text which is grounded in theories of communicative function, semiotics, and cognition offers

the potential to embrace both text and hypertext. If we additionally introduce the notion of genre as a solid frame of reference for comparison, there seems to be no overriding obstacle to comparing corpora of text and hypertext (for more on the problem of equation, cf. Moulthrop 1991).

Furthermore, it is obvious that current hypertexts are still strongly influenced by traditional norms for written text, in many cases lacking even basic features of hypertextuality. This is due to issues of intertextuality: many texts written and designed for traditional print media are (re)published on-line without major changes (pre-existing texts which are simply digitally distributed are known as *e-texts*). Further, hypertexts are currently produced by authors whose primary socialisation is typographical, and established paradigms are naturally not automatically left behind once they write for digital media. Our previous inter-medial research on genre conventions in traditional print and virtual environments has clearly shown that long-established textual conventions hardly disappear at once, but persist to an important extent in virtual environments (cf. Eckkrammer & Eder 2000).[2]

To come to a conclusion about the necessity of a modification of our traditional conception of *text*, in the light of hypertext, I assume that a wide semiotically, functionally, communicatively, and cognitively grounded conception of text still serves the purpose, and is particularly necessary for constrastive textology. Nevertheless, to a growing extent, it will be necessary for linguistics to integrate approaches from other disciplines (e.g. semiotics, cybernetics, cognitive sciences, and informatics), in order to provide adequate theoretical grounding for new forms of textual analysis.

4. Approaches to theories of inter-semiotic layering

Apart from the fact that hypertextual environments may transform the transmission of images into spatially and temporally delimited occurrences, for the purpose of contrastive textual analysis, I will take images here as integrative elements of (hyper)texts, intended to be displayed simultaneously with verbal elements, when activated by the reader. This is not to imply that a time-space axis will not play an important role in future analysis of hypertext. Indeed, as recently suggested by Miles (2000), who adopts schematic categories established by the movie semiotician Metz (1968) for hypertext syntagmas, a major challenge of hypertextual analyses will be to theoretically accommodate spatial or topographic and time-specific aspects of hypertext. However, frequently in current hypertexts, the elements linked to the time-space axis are seen to be largely related to tech-

nical constraints. On this basis, let us now turn to the model of inter-semiotic relationships, suggested by Hoek (1995) in the context of rhetorics.

4.1 The model proposed by Hoek (1995)

According to Hoek (1995), inter-semiotic layering falls into four dimensions. These lead from pure transposition – i.e. the transformation of image into verbal text or vice versa – to a degree of blending that does not permit a separation of the pictorial from the verbal – i.e. a fusion of both modes in syncretic[3] discourse. Table 1 illustrates the different types of inter-semiotic relationships in detail.[4]

In practical terms, transmedial relationships concern the kind of translation which Roman Jakobson termed inter-semiotic, i.e. the translation between different semiotic systems.[5] In the same way, we will have to take into account the possibility of intrapictorial translation (e.g. exchanging one drawing for another) and interpictorial translation (substituting a photograph for a drawing).

Within the simultaneous forms of verbal and pictorial layering, Hoek (1995: 77) distinguishes between juxtaposing, combining and fusing relationships, which differ particularly in the degree of separability (*separabilité*) and self-sufficiency (*autosuffisance*). The denser the level of *imbrication,* the more inseparable and less self-sufficient the different semiotic modes become. Hoek's pragmatically grounded classification is based on a clear distinction between the process of production and the process of reception. However, in hypertextual environments these processes may overlap, due to the fact that the reader may produce, in highly networked multi-semiotic hypertextual constructs, his individual textual reality with a unique pattern of inter-semiotic imbrication. Nevertheless, the analysis of 100 sample texts and hypertexts from the DIALAYMED corpus shows that the most frequent forms of layering remain juxtaposition and combination of verbal

Table 1. Types of inter-semiotic relationships (based on Hoek 1995)

Type: **Features:**	**Transmedial Relationship**	**Multimedial Discourse**	**Mixed Discourse**	**Syncretic Discourse**
Separability	Yes	Yes	Yes	No
Self-sufficiency	Yes	Yes	No	No
Process	Image is turned into verbal text, or vice versa	Pictorial and verbal elements coexist in a text without blending	Pictorial and verbal elements are combined in a text	Pictorial and verbal elements fuse or amalgamate in a text
Layering	Transposition	Juxtaposition	Combination	Fusion
Examples	Ekphrasis Photo-novel	Illustration Title	Poster/ad Comic	Visual poetry Calligram

elements with static images in a prearranged (hyper)textual setting which cannot be individually created.

4.2 Practical application of the model on sample (hyper)texts

As shown in the previous sections, meaning-making in contemporary (hyper)texts can be considered as a multi-faceted, dynamic process that happens once a reader accesses a text and begins to draw inferences according to their particular needs or interests. When people are concerned with a specific infectious disease, e.g. when planning a long-distance journey or when a friend or family member is affected by a particular illness, they frequently refer to texts such as medical encyclopaedia, pamphlets, and/or, increasingly, to hypertexts – such as those published by interest groups or official health agencies. Since infections usually represent complicated processes in the human body, popularising medical self-counselling genres may therefore refer to a variety of modes in order to explain these processes, e.g. how a virus or bacteria enters the body and unfolds harmful effects, how an infection can be avoided, or how to recognise and treat symptoms. The texts assembled in the DIALAYMED-corpus usually aim at explaining these issues to a lay audience. The contemporary medical self-counselling text is a multifunctional genre that incorporates specific practices of discourse, communication, production, interpretation, design and distribution which, within the multimodal theory of communication suggested by Kress and van Leeuwen (2001), determine the communicational ensemble and semiotic resources. According to Al-Sharief (1996: 11), the medical information leaflet, or medical self-counselling text, as we prefer to call the genre, performs the following tasks:

a. Providing a scientific background of the illness or health problem in question.
b. Preparing the reader/patient for the treatment by providing information about how the treatment will normally be carried out and what steps the doctor will take.
c. Persuading readers to stop unhealthy habits or take steps to reduce their harmfulness.
d. Giving practical advice that will help to avoid complications of the illness or will complement the treatment.
e. Arguing against misconceptions about the disease and/or its treatment.

This large variety of tasks can be achieved through a range of verbal and pictorial means. A look at selected sample texts illustrates the wide array of possibilities and also prompts the question of whether the intertwining of different modes furthers or inhibits these purposes. At this point it is important to note that the observations here will be limited to description of the (hyper)texts as 'products', since

Regimiẽto contra la peſte. Fecho
por el inſigne doctor Fernandalua
rez: medico de ſus altezas. Cathe-
dratico de ṕma en medicina eneſta
vniuerſidad de Salamanca.

Figure 1. Juxtaposing inter-semiotic layers in Alvarez Chanca: "Tratado nuevo, no menos útil que necesario, en que se declara de qué manera se ha de curar el mal de costado epidémico" (New treatise, not less useful than necessary, wherein the way to cure the epidemic breast disease is explained) from 1506 (reproduced on CD ROM, ADMYTE, p. 1)

observing receptive chains and outcomes of the reading process requires empirical analysis of real-world reading situations which is beyond the scope of this chapter.

Turning first to sample texts from the synchronic section of the DIALAYMED-corpus, it is relevant to note that medical self-counselling texts have included images since the advent of the genre, in the late 15th century. During the Middle Ages, when literacy was limited to very few people, pictorial information was of enormous value. In these early texts, one finds inter-semiotic layering of transmedial, multimedial, and blending types, similar to those in modern mixed and syncretic texts. The following figures illustrate the array of inter-semiotic layering in very early texts from the Spanish[6] section of the corpus.

Turning now to contemporary print and hypertext samples, it appears that inter-semiotic layering is strongly bound to the macrostructure of the (hyper)textual construct and also restricted to a limited number of texts. The verbal clearly dominates in both the traditional print and hypertext samples. At the same time, it is far from constituting an obligatory element in the genre. Functionally, the images predominantly show the relevant body parts, the ways in which the virus or bacteria is passed on, effects of the disease, test procedures and therapeutic measures taken by the doctor. In the vast majority of the texts analysed, the function of the images is purely illustrative, referring to something explicitly stated

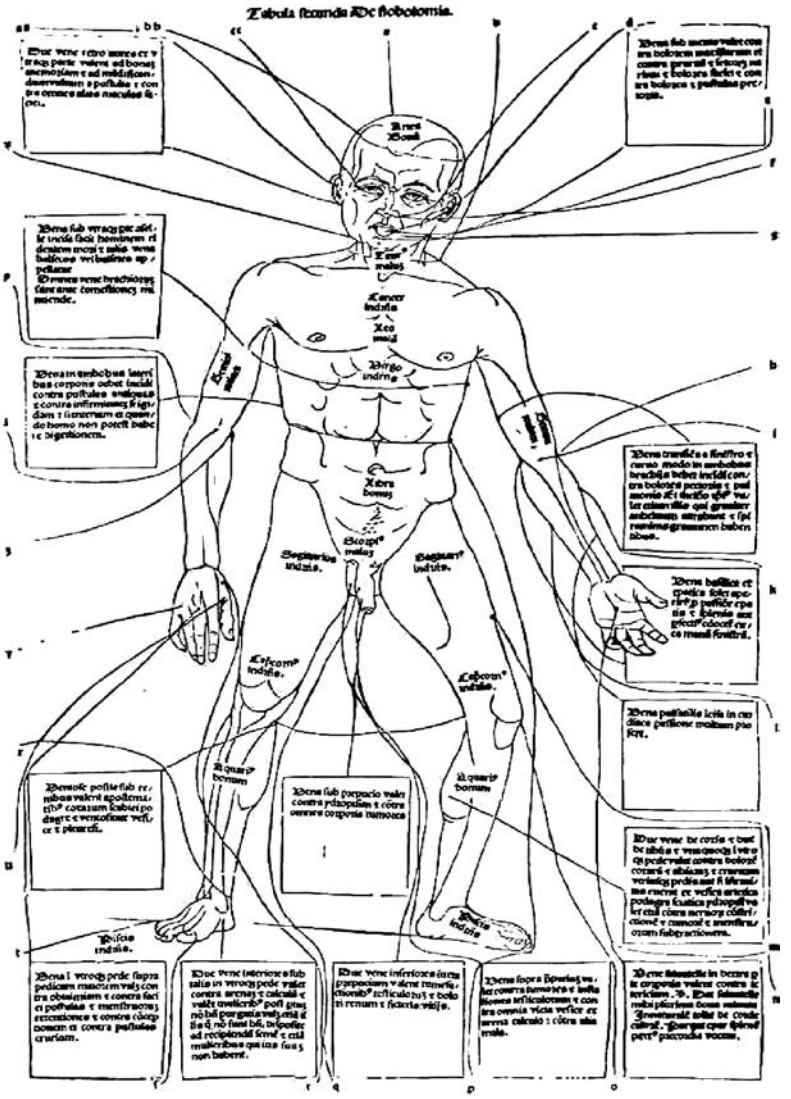
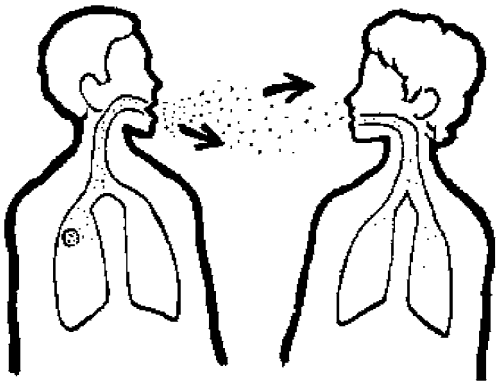


Figure 2. Combined inter-semiotic layering in Johannes von Ketham: "Compendio de la Humana Salud" (Manual of Human Health) from 1495 (reproduced in Herrera 1990: 57)

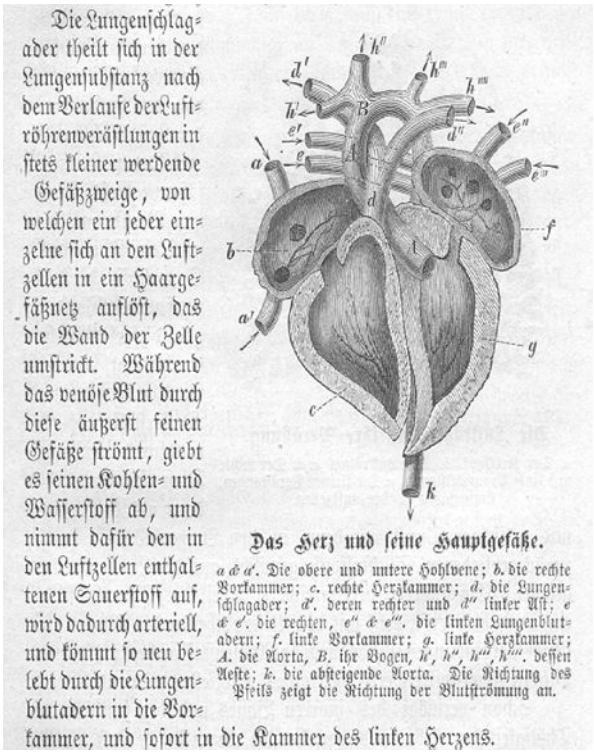
in the verbal text. Figures 3 and 4 from our data demonstrate this juxtaposing type of inter-semiotic layering.

What is of cross-linguistic and diachronic interest is the fact that only one of the 45 contemporary Spanish hypertexts in the corpus includes visual images. The making of meaning in multimodal text and hypertext is bound to many variables, particularly the underlying discourse, genre conventions, and cultural



TB germs spread though the air

Figure 3. Juxtaposed inter-semiotic layering which aims at unfolding a counselling function without a verbal element (from http://www.cpmc.columbia.edu/tbcpp/abouttb.html)



Die Lungenschlagader theilt sich in der Lungensubstanz nach dem Verlaufe der Luftröhrenverästlungen in stets kleiner werdende Gefäßzweige, von welchen ein jeder einzelne sich an den Luftzellen in ein Haargefäßnetz auflöst, das die Wand der Zelle umstrickt. Während das venöse Blut durch diese äußerst feinen Gefäße strömt, giebt es seinen Kohlen- und Wasserstoff ab, und nimmt dafür den in den Luftzellen enthaltenen Sauerstoff auf, wird dadurch arteriell, und kömmt so neu belebt durch die Lungenblutadern in die Vorkammer, und sofort in die Kammer des linken Herzens.

Das Herz und seine Hauptgefäße.

a & *a'*. Die obere und untere Hohlvene; *b*. die rechte Vorkammer; *c*. rechte Herzkammer; *d*. die Lungenschlagader; *d'*. deren rechter und *d''* linker Ast; *e* & *e'*. die rechten, *e''* & *e'''*. die linken Lungenblutadern; *f*. linke Vorkammer; *g*. linke Herzkammer; *A*. die Aorta, *B*. ihr Bogen, *h'*, *h''*, *h'''*, *h''''*. dessen Aeste; *k*. die absteigende Aorta. Die Richtung des Pfeils zeigt die Richtung der Blutströmung an.

Figure 4. Juxtaposed inter-semiotic layering with abstract explanatory devices in Günzburg: "Rathgeber für Brustschwache" (1873: 5)

context. Pictorial elements assist the construction of a continuity of meaning, if they are incorporated appropriately with respect to these variables. For example, the discursive habits and the degree of literacy of a culture determine whether prominence should be given to the verbal or to the pictorial code.

5. Challenges for the future: Intelligibility

Analysis of the DIALAYMED corpus has shown that (hyper)texts are often still much less multimodal than one would expect. This is particularly surprising in the case of the medical self-counselling genre, given its highly "mingled nature" and the "complex role" it must play in order to meet its objectives (cf. Al-Sharief 1996: 10). Reasons for this fact seem predominantly bound to costs: it is still more expensive to produce a multi-coloured brochure with well designed illustrations and/or other multimodal elements than a traditional black on white pamphlet limited to printed language. This difference does not, on the face of it, apply to hypertextual environments, so it may be surprising that – apart from inter-semiotic blending with pictorial elements – the degree of multimodality is not yet higher there than in print media. An important factor, however, is that technical devices and software for producing, accessing and retrieving medical information from, for example, video sequences or animated illustrations are not yet accessible to the vast majority. Moreover, the internet itself only represents a real option for information retrieval in the first world (cf. Global Internet Statistics 1997).

Nonetheless, multimodality research will have an important role to play in application to (hyper)texts such as these, particularly in the development of a dynamic and integrative theory of text intelligibility. This is because multimodality potentially impacts on all the dimensions which influence the comprehension and retention of text. I will proceed here on the basis of the four text intelligibility factors proposed by Schulz von Thun (1974) – (A) simplicity, (B) brevity, (C) stimulation and (D) textual organisation.

The first factor, simplicity, should not be seen only in terms of terminological and morpho-syntactic effortlessness (cf. Hohgräwe 1987). In some cases, a picture can certainly clarify a text, but may also make it more complex – for instance, if it is itself highly visually complex, incoherent in context, or included without cohesive co-text. In the case of brevity, a picture can definitely "paint a thousand words" and it is therefore possible to transpose a lot of verbal information into the pictorial mode, producing a more concise text. For example, in the one contemporary Spanish text which includes a pictorial element (photographs of syphilis symptoms), transposing this into language would certainly extend and complicate the verbal element. However, it should be remembered that the length of con-

temporary medical texts also reflects the enormous increase in medical knowledge during the past 100 years, and the relationship between length and clarity is not necessarily straightforward. The art of packing a lot of information into a text without scaring off the reader needs to take into account findings on the usability of hypertexts (cf. Nielsen 1994, 1997, 1999). In virtual environments readers literally scan through texts rather than read them as they would do with a printed version (cf. Krebs 1991 on content retention). This might imply the need for different strategies of inter-semiotic layering in medical self-counselling hypertexts, because the text needs to be even more concise and/or modularised than in traditional environments. Additionally, it is important to remember that the reader's expectations about what is appropriate to a genre (including its length) are highly relevant to the receptive process and may determine whether a text is read at all or immediately closed.

While multimodality research is needed to help illuminate the contribution of brevity and conciseness to making (hyper)text more understandable, there is an even clearer role for multimodality in the factor of stimulation. As observed in the DIALAYMED corpus, some stimulation can be achieved in the verbal mode through strategies such as direct producer-reader interaction (e.g. anticipated question-answer structures, cf. Gläser 1996; direct forms of instruction, etc.). However, well-designed multimodal textual elements certainly increase the reader's interest and curiosity, and appropriate inter-semiotic layering with pictorial elements can enhance stimulation and thereby foster text comprehension.

Finally, in textual organisation, which is the fourth of Schulz von Thun's factors for text intelligibility, multimodal elements are potentially seen to have important structural impact. Pictorial elements often guide the cognitive process during text reception in a number of ways. Medical self-counselling hypertexts, which draw on images for structure, are able to increase modularisation (see e.g. http://www.who.int/gtb/). Verbal elements can be broken into smaller portions with individual headings, a practice that boosts both redundancy between paragraphs and textual dynamism, two aspects linked to the success of popularising genres (cf. Pörksen 1986). Multimodal research into these kinds of potential could again support (hyper)text design for a better rate of information retention.

6. Conclusion

As this chapter is based on work in progress, it is more concerned with raising questions and contributing to discussion rather than providing definitive answers. Experience working with the DIALAYMED corpus has shown a number of areas where future multimodality research can contribute beneficially to specific

medical applications as well as more general (hyper)text design. There are also opportunities here for developing theories of multimodality and inter-semiotic layering to draw on research in areas such as text intelligibility (e.g. extending the findings of Schulz von Thun 1974), discourse comprehension (such as van Dijk & Kintsch 1983), web usability (e.g. Nielsen 1994) and more generally in the areas of corpus linguistics and contrastive textology.

Notes

1. This term is applied by Hoek (1995) in his French text. It originally denotes the form and way that roofing tiles are laid and passed on to a variety of metaphorical usages. In English the term belongs to geological terminology referring to the formation of an imbricate (layered) structure of sediments (shingle structure). The use may also be metaphorical (e.g. in politics).

2. Sometimes this continuity even leads to hybrid and, in our view, highly unacceptable texts in which the writer appears to be trapped between the norms of 'paper' genres and those of digital/hypertext genres.

3. The term syncretic, which is traditionally applied to refer to the combination of different forms of belief or religious practice, was adopted by contemporary semioticians to denote the fusion of two or more sign systems within one human activity, i.e. communication.

4. In Table 1, *discourse* is not applied in Foucault's sense (communicative practices which are distinctive), but, the way I understand the French original, it means *text* as communicative occurrence with a specific function.

5. The other two kinds of translation considered by Jakobson are *intralinguistic* translation, within one language (e.g. between English synonyms), and *interlinguistic* (and thus *intrasemiotic*) translation (e.g. substituting a French word for an English one).

6. Spanish is important in this period, as the Arabs of the Iberian Peninsula fostered the spread of ancient Greek and Roman medical knowledge (Hippocrates, Galen) and Arab advances (Avicenna, Rhazes).

References

Alexa, Melina & Zuell, Cornelia (1999). *A Review of Software for Text Analysis*. Mannheim: ZUMA (=GESIS Nachrichten Spezialband 5).

Al-Sharief, Sultan (1996). Interaction in Written Discourse. The Choices of Mood, Reference, and Modality in Medical Leaflets. University of Liverpool, unpublished doctoral dissertation.

Archivo Digital de Manuscritos y Textos Españoles (ADMYTE) I & II (1992). Madrid: Ministerio de Educación y Cultura & Micronet.

Bateman, John, Delin, Judy, & Henschel, Renate (this volume). "Multimodality and empiricism: preparing for a corpus-based approach to the study of multimodal meaning-making."

de Beaugrande, Robert-Alain & Dressler, Wolfgang U. (1981). *Einführung in die Textlinguistik*. Tübingen: Niemeyer.

van Dijk, Teun A. & Kintsch, Walter (1983). *Strategies of Discourse Comprehension*. Orlando: Academic Press.

Dressler, Wolfgang U. (1989). *Semiotische Parameter einer textlinguistischen Natürlichkeitstheorie*. Wien: Österreichische Akademie der Wissenschaften.

Dressler, Wolfgang U. (1999). "Semiotic preference structures in language." In G. F. Carr, W. Harbert, & L. Zhang (Eds.), *Interdigitations, Essays for Irmengard Rauch* (pp. 479–489). New York: Peter Lang.

Dressler, Wolfgang U. (2000). "Textlinguistik und Semiotik." In K. Brinker, G. Antos, W. Heinemann, & S. F. Sager (Eds.), *Text- und Gesprächslinguistik. Linguistics of Text and Conversation. Ein internationales Handbuch zeitgenössischer Forschung. An International Handbook of Contemporary Research*, Vol. 1 (pp. 762–772). Berlin: de Gruyter.

Dressler, Wolfgang U. & Eckkrammer, Eva M. (2001). "Functional explanation in contrastive textology." *Logos & Language, 2* (1), 25–43.

Eckkrammer, Eva M. & Eder, Hildegund M. (2000). *(Cyber)Diskurs zwischen Konvention und Revolution. Eine multilinguale textlinguistische Analyse von Gebrauchstextsorten im realen und virtuellen Raum.* Frankfurt am Main: Peter Lang.

Engebretsen, Martin (2000). "Hypernews and coherence." *Journal of Digital Information, 1* (7). http://jodi.ecs.soton.ac.uk/Articles/v01/i07/Engebretsen/. [19.12.2000]

Gläser, Rosemarie (1996). "Der implizite Dialog in populärwissenschaftlichen Texten im Deutschen und Englischen." In G. Budin (Ed.), *Multilingualism in Specialist Communication. Proceedings of the 10th European LSP Symposium Vienna, 29 Aug.–1 Sept., 1995*, Vol. I (pp. 751–770). Wien: Termnet.

Global Internet Statistics (1997 etc.). http://www.euro mktg.com/globstats

Günzburg, L. (1873). *Rathgeber für Brustschwache mit tuberkulöser Anlage, beginnender und ausgebildeter Lungenschwindsucht. Nach dem heutigen Standpunkte der Wissenschaft gemeinfasslich dargestellt. Zweite, gänzlich umgearbeitete und vermehrte Auflage. Mit 13 in den Text gedruckten Holzschnitten*. Wien – Pest – Leipzig: U. Hartleben's Verlag.

Herrera, Maria Teresa (1990). *Johannes de Ketham, Compendio de la humana salud*. Madrid: Arco Libros.

Hoek, Leo H. (1995). "La transposition intersémiotique pour une classification pragmatique". In L. H. Hoek & K. Meerhoff (Eds.), *Rhétorique et Image Textes en hommage à Á. Kibédi Varga* (pp. 65–80). Amsterdam: Rodopi.

Hohgräwe, Uwe (1987). *Verständlichkeit von Instruktionstexten und das Informationsverhalten von Arzneimittelverbrauchern.* Wuppertal: Fachbereich Gesellschaftswissenschaften der Bergischen Universität – Gesamthochschule Wuppertal (= Wuppertaler sozialwissenschaftliche Studien).

Kennedy, Graeme (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Krebs, Marlies (1991). Lesen oder "Navigieren im Hyperraum"?: eine empirische Studie zur Verständlichkeit von Hypertext im Vergleich zu linearem Text. University of Vienna, M.A. thesis.

Kress, Gunter & van Leeuwen, Theo (2001). *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold; New York: Oxford University Press.

Metz, Christian (1968). *Essais sur la signification au cinéma* /I/. Paris: Klincksieck.

Miles, Adrian (2000). "Hypertext syntagmas: cinematic narration with links." *Journal of Digital Information, 1* (7). http://jodi.ecs.soton.ac.uk/Articles/v01/i07/Miles

Moulthrop, Stuart (1991). "Beyond the Electronic Book: A Critique of Hypertext Rhetoric." In The Association of Computing Machinery (Ed.), *Proceedings of the Third ACM Conference on Hypertext* (pp. 291–298). San Antonio, TX: ACM.

Nielsen, Jakob (1994). *Usability Engineereing*. Cambridge, MA: AP Professional.

Nielsen, Jakob (1997). *Be Succinct!* (*Writing for the Web*). http://www.useit.com/alertbox/9703b.html

Nielsen, Jakob (1999). *Designing Web Usability: The Practice of Simplicity.* Indianapolis: New Rides Publishing.

Pörksen, Uwe (1986). *Deutsche Naturwissenschaftssprachen. Historische und kritische Studien.* Tübingen: Gunter Narr.

Schulz von Thun, Friedemann (1974). "Verständlichkeit von Informationstexten: Messung, Verbesserung und Validierung." *Zeitschrift für Sozialpsychologie, 5*, 124–132.

Storrer, Angelika (2000). "Was ist "hyper" am Hypertext?" In W. Kallmeyer (Ed.), *Sprache und Neue Medien* (pp. 222–249). Berlin and New York: de Gruyter.

Chapter 12

On the multimodality of interpreting in medical briefings for informed consent

Using diagrams to impart knowledge*

Kristin Bührig
University of Hamburg, Germany

This chapter deals with multimodality in untrained interpreting in an institutional context: mediated doctor-patient communication in German hospitals. The focus is on investigating a particular type of hospital interaction with a patient, 'the briefing for informed consent', and more specifically in it, the doctor's and the non-trained interpreter's use of a diagram in the source language (German) and the target language (Portuguese). The analysis will show how the different use of the diagram affects the respective speech action patterns and indicates what the implications are for imparting knowledge to the patient.

1. Interpreting in hospitals

In German hospitals today, when German-speaking doctors meet patients who, in either their own estimation or that of the medical personnel, do not possess sufficient knowledge of German to hold a discussion, the task of interpreting is often given to family members, friends, or members of the care or cleaning personnel. Interpreting in hospital has the aim of enabling understanding between a doctor and a patient. It is true that even monolingual doctor-patient communication shows deficits in terms of linguistic cooperation, resulting, among other things, from barriers due to specialist language. (The literature on comprehension problems in doctor-patient communication is extremely extensive, and a discussion of the issues is beyond the scope of this chapter. For an overview, see e.g. Löning & Rehbein 1993; Todd & Fisher 1993; Redder & Wiese 1994.) It could be expected, however, that interpretation would remove at least some of the difficulties which result from the barriers between two languages. However, those persons commissioned with interpreting perform their task without being financially rewarded or prepared for it. The use of an interpreter is also generally ad hoc and without prior consultation with the medical personnel.

The project "Interpreting in Hospitals"[1] aims to investigate what the communicative demands are which these interpreters face and what effect interpreting has on doctor-patient communication. The focus here is on a particular type of hospital interaction. In Germany, prior to diagnostic and therapeutic operations, patients are informed about the procedure for the operation as well as any possible complications, side effects or consequences. In hospitals and medical training contexts, these discourses are known as 'briefings for informed consent'. This chapter presents an analysis of an interpreted briefing for informed consent which includes visuals in the form of a diagram. The focus will here be on the impact of multimodality on this kind of hospital interpreting. The next section outlines the theoretical framework and analytical procedure.

2. Interpreting and multimodality

The task of interpreting is widely represented in the research literature as the overcoming of barriers between two languages in spoken discourse. This is an important point of difference between interpreting and translation (for an overview of translation definitions and terminology, see Pöchhacker 2000). However, apart from this differentiation between translating and interpreting, the special medial and modal characteristics of interpreting have so far been largely neglected, both in theory and in the practice of interpreting. For example, simultaneous conference interpreting is very often based on the written notes of the speaker and thus shows parallels with on-sight translation (Bührig 1999). In consecutive interpreting, the interpreter frequently makes written notes to which s/he then refers to (Durlanik 2001), and in institutional contexts, written documents and diagrams are often included in communication by the representatives of the institution. Research in a range of linguistic fields has shown that linguistic action is bound to its context and that it will therefore differ in its concrete form according to which contextual elements – i.e. 'environments' (Bührig & ten Thije 2001) – are necessary in the understanding process. For example, a conversation referring to the immediate situational context of the 'Wahrnehmungsraum' (*perception space*: Ehlich 1979, following Bühler 1934) shows different features from a discussion on the telephone, in which the same perception space is not available for both linguistic actors (on the use of the perception space in communication via computer, see e.g. Brinkschulte & Grießhaber 1999). There is wide consensus that written and spoken modes are associated with different kinds of information processing (e.g. Chafe 1982, 1992; Ehlich 1994a; Halliday 1989; Koch & Österreicher 1985; Raible 1996; Rehbein 2000, 2001), and, more generally, that the medium of communication affects the linguistic form.

The first step in investigating the impact of multimodality on hospital interpreting and doctor-patient communication is a comparison of source and target language expressions in the interaction. This comparison, following the theoretical framework of 'functional pragmatics', is made from the hearer's perspective, so as to reconstruct her/his processes of understanding (Ehlich 1991; Rehbein 2001). The reconstruction also clarifies which dimensions of the source language action are 'reproduced' in the target language (Bührig & Rehbein 2000) – for example, whether target and source language actions differ in their modal or medial features. The texts which have been analysed in the current project are interpreted 'briefings for informed consent' which include diagrams. Excerpts of the discourse show how diagrams are linked to both the linguistic action of the doctor and to the imparting of knowledge to the patient. The second step in the investigation is to reconstruct how, and to what extent, the interpreter uses the visually perceptible environment and what effect the visuals have on conveying knowledge in the target language.

At this point it is important to note that the difference between interpreting and translation implies an additional complication for analysis; it is not always the case in interpreting situations that the main participants speak and understand only their own language. Participants in the discussion may understand at least partially what is said in the other language. In these cases, interpreting is said to be 'transparent' (Müller 1989: 716). The interactions which we are researching are primarily based on transparent interpreting, since, although not considered sufficient for communicating with the doctor, each patient has at least some knowledge of German.

3. The use of diagrams in medical briefings for informed consent

Briefings for informed consent have certain communicative components in common (Meyer 2000): announcement of the operation, description of the operation (procedure and aim, describing the organs involved), discussion of the complications (listing and illustrating complications and assessing the risk), and checking the need for further briefing. According to the type of operation, each element is given a different weighing or discursive treatment. In the briefing, the use of diagrams (often included on an information sheet) may take place either when knowledge about the purpose of the operation or its procedure is given. In the analysed data, there are two ways of including diagrams: some doctors make the diagrams (and particularly the illustrations of the relevant organs) the starting point for their description; other doctors first describe the operation and then refer to

the illustrations, without using them again during the briefing. In the following section, an example of the first kind of diagram use will be looked at more closely.

4. Briefing for a gallstone operation

The discussion and the analysis which follow are from a briefing for a Portuguese patient's informed consent to an operation to examine her bile ducts and pancreas. The patient is about 55-years-old and has lived in Germany for many years. Although the hospital staff has defined her German as good, all important discussions are interpreted. In the present discussion, the patient's daughter is interpreting. She is about 17-years-old and was born and brought up in Germany.

The patient was taken to hospital with a stomach-ache. She had previously undergone an operation in another hospital to remove her gallstones. Before this conversation with the doctor her bile ducts were x-rayed. The new examination is necessary from the doctor's point of view, because he suspects that a gallstone in the patient has moved through the bile ducts to the papilla and is causing a blockage of gall and digestion juices there, and has resulted in an inflammation of the bile ducts and the pancreas. If this suspicion is confirmed by the examination, the doctor plans to remove the gall-bladder. The examination, which is the subject of the present discussion, is therefore not necessarily the last invasive operation to be carried out on the patient in this hospital.

The doctor's explanation of the suspected diagnosis is presented in (1).

(1) Also es geht darum, dass •• wir ja gefunden haben, dass in der Gallenblase Steine sind. ((1second pause)) Ja? •• Und wir vermuten, dass einer von diesen Steinen in die Gallenwege abgerutscht ist und eine akute Entzündung in den Gallenwegen und in der Bauchspeicheldrüse gemacht hat.
So, this is about the •• fact that we have found that there are stones in the gallbladder. ((1 second pause)) Okay? •• And we suspect that one of these stones has slid down into the bile ducts and • has caused an acute inflammation in the bile ducts and in the pancreas.

In response to the patient's query "Und wieso kommt lange Schmerz?" (*And why is there so much pain?*) the doctor first gives some information about the organs affected by the illness, and then explains the suspected diagnosis again. To begin with, we shall look at the first step in imparting medical knowledge (i.e. introducing the organs), the use of the diagram, and the interpretation into Portuguese.

4.1 Introducing the organs in the source language

Before the doctor brings his knowledge into the discourse, he first prepares the use of the diagram, in that he asks the patient to look at it, and justifies its use (2).

(2) Ich kann Ihnen das mal zeigen. Sie können ja hier mit gucken. Dass Sie das gleich verstehen.
I can show it to you. You can look at this here with me. So that you understand it straight away.

In the comments which follow, "Das ist die *Leber*" (*That is the **liver***) and "Das ist die Gallenblase" (*That is the gall-bladder*), the doctor orients the patient's attention to the visually illustrated organs involved in her illness, through the deictic "das" (*that*) (for further discussion of such deictics, see Ehlich 1979, 1982, 1983). As a starting point for building knowledge, the doctor thus uses the patient's visual perception, directing her focus to the organs which he then gives names. Both statements, therefore, show the form "das" (deictic reference), "ist" (*is*) (copula), "die Leber" (*the liver*) / "die" Gallenblase" (*the gall-bladder*) (definite article + noun). Ehlich (1994b) defines this structure, in conjunction with an image, as an 'ostensive definition'. With the introduction of new objects to a hearer's consciousness, ostensive definitions create a link between a word and a graphically represented element of reality. The deixis realised by "das" is purely for the orientation of the hearer to the object (ostension), to link it to an element in the symbol field[2] of language, i.e. the noun (cf. Ehlich 1991; Rehbein 1998). Such ostensive definitions can, in my view, be understood to initiate processes of building knowledge, as they allow the identification of elements about which the hearer has no prior knowledge. They form the perceptual basis of knowledge verbalised by a speaker, and so guarantee its sensory certainty through the orientation towards the diagram. Knowledge established in this way can then be enriched with further knowledge elements. This is exactly what happens in further discussion between the doctor and the patient. In a subsequent step, again using a deictic expression, "da" (*there*), the doctor first indicates the gallstones, then mentions the function of the gall-bladder and the transport of the bile into the intestines (3).

(3) Da sitzen die Steine drin, nech? Und • das is • ein Speicher für die Galle. Und de/ ••• die Galle wird • über den Gallengang •• in den Darm abgegeben.
The stones are sitting in there, aren't they?. And • that's • where the bile is stored. And the/ ••• the bile is • passed via the bile ducts •• into the intestines.

In addition, with the use of the symbol field expression "Speicher" (*storage place*), the prepositional phrase "über den Gallengang" (*via the bile ducts*), and the passive, the doctor verbalises the functional interrelation in the digestion process between the organs shown and named (the gall-bladder and the bile ducts).

These explanations by the doctor are characterised by clear, short statements, which ensure a successive build-up of knowledge. This verbalised knowledge also has a characteristic structure across the utterances, which can be traced back to the use of the diagram. That is, at the outset, the doctor gives a name to each visually perceivable object with ostensive definitions, through the use of deictic expressions and in locating the gallstones. With the help of a combination of the introduced symbol field expressions and definite articles, this knowledge then becomes the 'theme' (the first element) of the following utterances. In this way, there is an increased connection between the discourse knowledge that is already established and the newly verbalised knowledge that is the 'rhematic' (following) element in the next statement.

4.2 The organs in the target language

The next step is to compare these linguistic actions of the doctor with the actions of the interpreter in the target language, Portuguese. Table 1 is a transcript of the relevant part of the discussion, organised by speaker. The doctor's utterances are grouped into the column labelled "A", the utterances of the patient's daughter, who is interpreting, are headed "D" and the patient's utterances are in the column labelled "P". An English gloss is given in italics beneath each utterance. Text in double parentheses shows the length of pauses in seconds, e.g. "((1.2 s))". A dot indicates a pause of a quarter of a second, two dots a half second, e.g. "• *aren't they*?". Bold text indicates intonationally emphasised parts of an utterance, e.g. "Auf dem **Bild**".

The first contrast that is apparent between the doctor's utterances and those of the interpreter is that the daughter does not convey the request that the patient should direct her attention to the information sheet or the diagram in the target language. She begins her interpreting in Segment 36, with the doctor's statement about the transport of the bile into the intestines (Segment 34). However, she does not finish this statement, but instead first indicates – probably visually – the gall-bladder to the patient: "Äh • aqui é o coisu, ja?" (*Ah • here's the thing, right*?) (Segment 37).

Let us look more closely at the inclusion of the diagram in the interpreter's Portuguese utterances. Like the doctor, she uses a deictic expression, "aqui" (*here*). However, "aqui" refers only to location, and not, like the German "das" (*that*), to an object in the perceptible space. Perhaps the interpreter uses "aqui" to direct the patient's attention back to the place in the diagram to which the doctor had referred in his German explanation. That is, it is conceivable that the patient's gaze had already followed as the doctor pointed lower than the gall-bladder in the diagram, to trace the transport of bile into the intestines (Segment 34). In this case,

Table 1. Transcript excerpt showing the organs

A	D	P
		(23) Und wieso kommt lange Schmerz? *Und why is there so much pain?*
(24) Ich kann Ihnen das mal zeigen. *I can show it to you.*		
(25) Sie können ja hier mit gucken. *You can look at this here with me.*		
(26) Auf dem **Bild**, was ich Ihnen mitge- bracht habe. *On the picture that I brought for you.*		
(27) Haben Sie ne • Brille da, oder? *You have glasses there, don't you?*		
(28) ((4,7s)) Dass Sie das gleich verstehen. *((4.7s)) So that you understand it straight away.*		
(29) ((2,8s)) Das ist die **Leber**. *((2.8s)) That is the liver.*		
		(30) •• Hmhm
(31) ((1,2)) Das ist die Gallen**bla**se. *((1.2s)) That is the gall-bladder.*		
	(36) Also vai pra... *So that goes to...*	
(32) ((1,3s)) Da sitzen die Steine drin, • nech? *((1.3s)) The stones are sitting in there, • aren't they*?		
(33) Und •• das is •• ein Speicher für die Galle. *And •• that's •• where the bile is stored.*	(37) Äh • aqui é o coisu, ja? *Ah • here's the thing, right*?	
	(38) Tu dizes que é o veneno dos coelhos. *You say,it's the rabbit poison.*	

Table 1. (*continued*)

A	D	P
		(39) Bom *Well.*
(34) Und de/ ••• die Galle wird • über den Gallengang •• in den Darm gegeben. *And th/ •• the bile is • passed via the bile ducts •• into the intestines.*	(41) Vai. *It goes*	
	(42) E depois vai para os intenstinos. *And then it goes to the intestines.*	
(35) Normalerweise. *Normally.*		

the use of "aqui" could be traced to the 'internally dilated speech situation' (Bührig & Rehbein 2000), which is responsible for the lack of simultaneity in consecutive interpreting.

A further difference between source and target expression is that, in the place of the doctor's ostensive definitions, the interpreter gives no technical names to the organs. Rather, in Segment 38, she initially vaguely qualifies the element shown as the subject, with "o coisu" (*the thing*), and then adapts the relevant knowledge to her mother's prior knowledge: "Tu dizes que é o veneno dos coelhos." (*You say it's the rabbit poison.*).

After the mother has signalled her understanding (Segment 39), the daughter again speaks about the transport of the bile into the intestines (Segments 40 and 41), without mentioning the gallstones. The interpreter does not mention the function of the gall-bladder in the target language, nor does she convey the Portuguese symbol field expressions for all the organs (liver, gall-bladder, bile ducts, bile). Only the "intestines" (*intestinos*) are named.

Comparable to the build up of knowledge by the doctor, the interpreter also tries to use a successive procedure. However, a glance at the knowledge organisation in the target language shows that the daughter's statements do not cover several topics, but only one, namely the knowledge verbalised with "o coisu" (*the thing*) or "o veneno do coelhos" (*rabbit poison*), with which the prior knowledge of the patient is linked. The interpreter continues this topic with the help of the 3rd person nominative morphology of the finite verb "vai" (*go*).

In the absence of a visual record, we can only guess to what extent the interpreter uses the diagram. However, the intonation patterns suggest that the

interpreter makes accompanying gestures with her statement showing the route of the bile to the intestines. In Segment 41 she verbalises the destination of the gallstone visible on the diagram with "para os intestinos" (*to the intestines*). In this following utterance, the interpreter uses the verb "vai" (*it goes*) once again so that the functional interrelation verbalised by the doctor between the gall-bladder, the gallstone, and the intestines assumes the character of an iterative procedure in the target language, which she sub-divides temporally in Segment 41 with "e depois" (*and then*).

4.3 The suspected diagnosis in the source language

In Segment 43, based on the verbalised knowledge about the relationship between the gall-bladder, the gallstone, and the bile ducts, the doctor begins to speak about a disruption of this functional interrelation, as has presumably happened with the patient (4).

(4) Und wenn jetzt ein Stein dort in der Gallenblase sitzt, kann es eben passieren, dass der Stein ••• auch in die Gallenwege abrutscht.
And if there is a stone sitting there in the gall-bladder now, there is even a chance that the stone ••• could also slide down into the bile ducts.

In the subordinated clause of his statement, the doctor links the hearer's perception, through the deictic expression "dort" (*there*) and the prepositional phrase "in der Gallenblase" (*in the gall-bladder*), with the knowledge established previously about the stones in the gall-bladder in Segments 31 and 32. In the main clause, using the modal verb "kann" (*can*), he then expresses the possibility that a gallstone has found its way from the gall-bladder into the bile ducts.

If we look at this statement more closely, the knowledge verbalised by the doctor shows traces of its generation. This expression has a hypothetical character, which the doctor verbalises with the expression "wenn" (*if*), assigning a purely linguistic reality to the knowledge verbalised in the subordinated clause (cf. Redder 1990: 265). This hypothetical character is also created with the use of the modal verb "kann" (*can*) (cf. Ehlich & Rehbein 1972; Redder 1984). However, at the same time, the doctor qualifies the knowledge expressed in the main clause with the particle "eben" (*even*), indicating it to be based on prior knowledge, so that the possibility of a gallstone moving into the bile ducts assumes the quality of a 'practical conclusion' that is derived from specialist medical knowledge (Ehlich & Rehbein 1979). According to Ehlich (1984), the combination of "wenn" with the modal verb "kann", can be understood as an attempt by the doctor to set in motion a 'knowledge-based imagination activity' on the part of the hearer, to guarantee comprehension of the suspected medical diagnosis.

4.4 The suspected diagnosis in the target language

Table 2 provides the transcript for the subsequent part of the briefing covering the patient's suspected diagnosis. (Transcription notation is as for Table 1.)

In Segment 44, the interpreter makes the information expressed in the doctor's main clause the central point of the target language statement, using the matrix construction "pode ser" (*maybe*). Like the doctor, she thus qualifies the knowledge in question as a possibility, but does not transfer into the target language the full knowledge verbalised by the doctor in the source language. Whereas the doctor made an effort to use the discourse knowledge which has already been built up for further reception (cf. Rehbein et al. 1995), the interpreter starts the speech situation anew. There is no linguistic link to the previous target-language statements, and thus the statement is not grounded in the discourse knowledge.

In the source language, the suspected diagnosis concerns the possible sliding of a gallstone. In the target language, however, it is concerned with the possibility of a stone perhaps being located in a certain place. This localisation is taken over by the interpreter with the aid of the diagram, on which she focuses the attention of the patient by using the deictic expression "lá" (*there*). While the doctor, using the local distant deixis "dort" (*there*), creates a perceptual basis for the reception of the *thematic,* first part of his statement in the previous relative clause, the interpreter uses the perception of the patient for the *rhematic*, final part of the statement. (For

Table 2. Transcript excerpt showing the suspected diagnosis

	A		D	P
(43)	Und wenn jetzt ein Stein dort in der Gallenblase sitzt, kann es eben passieren, dass der **Stein** ••• **auch** in die Gallenwege abrutscht. *And if there is a stone sitting there in the gall-bladder now, there is a chance that the stone ••• could also slide down into the bile ducts.*	(44)	Pode ser que é uma pedra também tinha lá. *Maybe there was also a stone there.*	
(45)	Nich? *Right?*			
(46)	• Das macht Schmerzen. • *That causes pain.*			
(47)	••• Das tut weh. ••• *That hurts.*			
				(48) Hmhm

a discussion of the relevance of the position of linguistic elements for discourse analysis, see Rehbein 1992, 1995.)

5. Conclusion

In her question in Segment 23, "Und wieso kommt lange Schmerz?" (*And why is there so much pain?*), the patient expresses a deficit of knowledge which – and this is clear through the initial "und" (*and*) – affects the link between the suspected diagnosis, expressed in Segment 18 by the doctor, and her pain. The doctor fills this deficit of knowledge with certain knowledge elements, which clarify the suspected diagnosis through the functional interrelation between the relevant digestive organs. Through these utterances, the doctor realises the speech action pattern of 'explaining' (Ehlich & Rehbein 1986; Rehbein 1985, 1994; Bührig 1996).

In her target-language action, the interpreter does not, however, represent the functional interrelation between the organs. Instead, with the expression "veneno dos coelhos" (*rabbit poison*), she uses one of the patient's own formulations to describe the bile. Thereafter, the prior knowledge of the patient updated in this way becomes the topic of the statements in Portuguese, which are based on the doctor's German statements about the transport of the bile to the intestines. With "vai" (*it goes*), the interpreter names only one external aspect of this procedure, the movement. Her statements can therefore be best understood as 'describing' (Rehbein 1984). The subject of her description – in comparison to the source-language actions – is reduced to the transport of the bile to the intestines. Thus, she cannot link the suspected diagnosis to the knowledge already verbalised in the target language, but, in contrast to the doctor, makes a new start and then uses the diagram in the *rhematic* part of her expression, mentioning a gallstone with the deictic expression "lá" (*there*) in the perception space. So, from an overall comparison of the source-language and target-language utterances it can be seen that the doctor and the interpreter are indeed using two different 'linguistic action patterns'.

From the point of view of the visual elements, the analysis has shown that the medical doctor and the non-trained interpreter also use different multimodal action patterns. The doctor uses the diagram to systematically build up the patient's knowledge, according to his own professional medical knowledge, through the speech action pattern of 'explaining'. First, he uses the diagram for ostensive definitions. Later, he uses the diagram and deictic expressions to refocus *thematic* elements of knowledge. By contrast, the interpreter works in a more local way. She links the diagram to new knowledge and the *rhematic* parts of her statements.

What causes the interpreter to reproduce the doctor's statements only partially? One factor may be that the transparent interpreting situation is at fault. Due

to the – albeit hesitant – hearer-signal of the patient in Segment 30, ("Hmhm"), the interpreter may have assumed that her mother had understood the ostensive definitions and that therefore they did not need to be interpreted. On the other hand, the unspecific naming of the gallstone with "o coisu" (*the thing*) in Segment 37 also raises the suspicion that the interpreter does not herself know what the respective organs are called in Portuguese. A third possibility is that the interpreter is trying to adapt the knowledge verbalised by the doctor to suit her mother's prior knowledge, with her use of "veneno dos coelhos" (*rabbit poison*) for bile. Whatever the causes, the doctor and the interpreter noticeably make use of the diagram in different ways, and employ different systems of imparting knowledge.

Transcripts of a later briefing for informed consent with this patient show that she is not in fact able to use the appropriate knowledge about her digestive organs. Leaving aside the issue of training in medical terms or interpreting, it is clear that untrained interpreters, who may lack skills or have different orientations to the build-up of knowledge by integrating language and other modes, such as visuals, may, without preparation, very often be quite unable to overcome the language barriers between doctors and patients in briefings for informed consent.

Notes

* An earlier version of this chapter was presented at the 32nd meeting of the *Gesellschaft für Angewandte Linguistik* (GAL, Passau, 27–29th September, 2001). I would like to thank Jutta Fienemann, Katharina Meng, Bernd Meyer and Jan ten Thije for their helpful comments on this presentation. For help with the English version I am indebted to Ann Helin Langridge.

1. Others currently working on this project are Bernd Meyer and Erkan Özdil. The financial support, since 1999, by the Deutsche Forschungsgemeinschaft (DFG) of "Interpreting in Hospitals" is gratefully acknowledged.

2. In the context of functional-pragmatic discourse analysis, following Bühler (1934), who differentiates between the 'Zeigfeld' (*descriptive field*) of a language and the 'Symbolfeld' (*symbol field*), nouns and verb and adverbial stems are analysed as elements of the 'symbol field', with which the speaker directs the listener to complete an 'appellative procedure' (on the concept of the 'procedure', see e.g. Ehlich 1991, 1993).

References

Brinkschulte, Melanie & Grießhaber, Wilhelm (1999). "Gestisches Sprechen. Sprechen vor dem Computer." In *Osnabrücker Beiträge zur Sprachtheorie* (*OBST*), *60*, 33–50.

Bühler, Karl (1934/1982). *Sprachtheorie. Zur Darstellungsfunktion der Sprache*. München: Fink.

Bührig, Kristin (1996). *Reformulierende Handlungen. Zur Analyse sprachlicher Adaptierungsprozesse in institutioneller Kommunikation.* Tübingen: Gunter Narr.

Bührig, Kristin (1999). "Konsekutives Übersetzen Englisch-Deutsch." In H. Gerzymisch-Arbogast, D. Gile, J. House, & A. Rothkegel (Eds.), *Wege der Übersetzungs- und Dolmetschforschung* (pp. 241–266). Tübingen: Gunter Narr.

Bührig, Kristin & Rehbein, Jochen (2000). "Reproduzierendes Handeln. Übersetzen, simultanes und konsekutives Dolmetschen im diskursanalytischen Vergleich." *Arbeiten zur Mehrsprachigkeit, Folge B*, 6/2000.

Bührig, Kristin & ten Thije, Jan (in press). "Diskurspragmatische Beschreibung." Forthcoming in U. Ammon, N. Dittmar, & K. Mattheier (Eds.), *Sociolinguistics / Soziolinguistik. An International Handbook of the Science of Language and Society. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. Berlin: de Gruyter.

Chafe, Wallace L. (1982). "Integration and involvement in speaking, writing and oral literature." In D. Tannen (Ed.), *Spoken and Written Language: Exploring Orality and Literacy* (pp. 35–53). Norwood: Ablex.

Chafe, Wallace L. (1992). "Information flow in speaking and writing." In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The Linguistics of Literacy* (pp. 17–29). Amsterdam, Philadelphia: Benjamins.

Durlanik, Latif M. (2001). *Notizen und verbales Planen. Diskursanalytische Untersuchungen zum Konsekutivdolmetschen Deutsch / Türkisch.* Münster: Waxmann.

Ehlich, Konrad (1979). *Verwendungen der Deixis beim sprachlichen Handeln. Linguistisch-philologische Untersuchungen zum hebräischen deiktischen System*. Frankfurt: Peter Lang.

Ehlich, Konrad (1982). "Deiktische und phorische Prozeduren beim literarischen Erzählen." In E. Lämmert (Ed.), *Erzählforschung* (pp. 112–129). Stuttgart: Metzler.

Ehlich, Konrad (1983). "Deixis und Anapher." In G. Rauh (Ed.), *Essays on Deixis* (pp. 79–97). Tübingen: Gunter Narr.

Ehlich, Konrad (1984). "Handlungsstruktur und Erzählstruktur. Zu einigen Kennzeichen der Weiterentwicklung von Erzählanfängen." In K. Ehlich (Ed.), *Erzählen in der Schule* (pp. 126–175). Tübingen: Gunter Narr.

Ehlich, Konrad (1991). "Funktional-pragmatische Diskursanalyse." In D. Flader (Ed.), *Verbale Interaktion. Studien zur Empirie und Methodologie der Pragmatik* (pp. 127–143). Stuttgart: Metzler.

Ehlich, Konrad (1993). "Sprachliche Prozeduren in der Arzt-Patienten-Kommunikation." In P. Löning & J. Rehbein (Eds.), *Arzt-Patienten-Kommunikation. Analysen zu interdisziplinären Problemen des medizinischen Diskurses* (pp. 67–90). Berlin and New York: de Gruyter.

Ehlich, Konrad (1994a). "Funktion und Struktur schriftlicher Kommunikation." In H. Günther & L. Otto (Eds.), *Schrift und Schriftlichkeit. Writing and its Use. 1. Halbband* (pp. 18–41). Berlin and New York: de Gruyter.

Ehlich, Konrad (1994b). "Verweisungen und Kohärenz in Bedienungsanleitungen. Einige Aspekte der Verständlichkeit von Texten." In K. Ehlich, C. Noack, & S. Scheiter (Eds.), *Instruktion durch Text und Diskurs. Zur Linguistik 'technischer Texte'* (pp. 116–149). Opladen: Westdeutscher Verlag.

Ehlich, Konrad & Rehbein, Jochen (1972). "Einige Interrelationen von Modalverben." In D. Wunderlich (Ed.), *Linguistische Pragmatik* (pp. 318–340). Frankfurt am Main: Athenäum.

Ehlich, Konrad & Rehbein, Jochen (1979). "Sprachliche Handlungsmuster." In H.-G. Soeffner (Ed.), *Interpretative Verfahren in den Sozial- und Textwissenschaften* (pp. 243–274). Stuttgart: Metzler.

Ehlich, Konrad & Rehbein, Jochen (1986). *Muster und Institution. Untersuchungen zur schulischen Kommunikation.* Tübingen: Gunter Narr.

Halliday, M. A. K. (1989). *Spoken and Written Language*. Oxford: Oxford University Press.
Koch, Peter & Österreicher, Wulf (1985). "Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte." *Romanistisches Jahrbuch, 36*, 15–43.
Löning, Petra & Rehbein, Jochen (1993). *Arzt-Patienten-Kommunikation. Analysen zu interdisziplinären Problemen des medizinischen Diskurses.* Berlin and New York: de Gruyter.
Meyer, Bernd (2000). "Medizinische Aufklärungsgespräche. Struktur und Zwecksetzung aus diskursanalytischer Sicht." *Arbeiten zur Mehrsprachigkeit. Folge B,* 8/2000.
Müller, Frank (1989). "Translation in Bilingual Conversation: Pragmatic Aspects of Translatory Interaction." *Journal of Pragmatics, 13*, 713–739.
Pöchhacker, Franz (2000). *Dolmetschen. Konzeptuelle Grundlagen und deskriptive Untersuchungen*. Tübingen: Stauffenburg.
Raible, Wolfgang (1996). "Orality and literacy. On their medial and conceptual aspects." In D. Scheunemann (Ed.), *Orality, Literacy and Modern Media* (pp. 17–26). Columbia: Camden House.
Redder, Angelika (1984). *Modalverben im Unterrichtsdiskurs. Pragmatik der Modalverben am Beispiel eines institutionellen Diskurses*. Tübingen: Niemeyer.
Redder, Angelika (1990). *Grammatiktheorie und sprachliches Handeln:* ***denn*** *und* ***da****.* Tübingen: Niemeyer.
Redder, Angelika & Wiese, Ingrid (1994). *Medizinische Kommunikation. Diskurspraxis, Diskursethik, Diskursanalyse.* Opladen: Westdeutscher Verlag.
Rehbein, Jochen (1984). "Beschreiben, Berichten und Erzählen." In K. Ehlich (Ed.), *Erzählen in der Schule* (pp. 67–124). Tübingen: Gunter Narr.
Rehbein, Jochen (1985). "Medizinische Beratung türkischer Eltern." In J. Rehbein (Ed.), *Interkulturelle Kommunikation* (pp. 349–419). Tübingen: Gunter Narr.
Rehbein, Jochen (1992). "Zur Wortstellung im komplexen deutschen Satz." In L. Hoffmann (Ed.), *Deutsche Syntax. Ansichten und Aussichten* (pp. 523–574). Berlin and New York: de Gruyter
Rehbein, Jochen (1994). "Rejective proposals. Semi-professional speech and clients' varieties in intercultural doctor-patient-communication." *Multilingua, 13* (1/2), 83–130.
Rehbein, Jochen (1995). "Grammatik kontrastiv – am Beispiel von Problemen mit der Stellung finiter Elemente." *Jahrbuch Deutsch als Fremdsprache, 21*, 265–292.
Rehbein, Jochen (1998). "Die Verwendung von Institutionensprache in Ämtern und Behörden." In L. Hoffmann, H. Kalverkämper, & E. H. Wiegand (Eds.), *Fachsprachen. Handbücher zur Sprach- und Kommunikationswissenschaft 14.1* (pp. 689–709). Berlin and New York: de Gruyter.
Rehbein, Jochen (2000). "Prolegomena zu Untersuchungen von Diskurs, Text, Oralität und Literalität unter dem Aspekt mehrsprachiger Kommunikation." In B. Meyer & N. Toufexis (Eds.), *Text / Diskurs, Oralität / Literalität unter dem Aspekt mehrsprachiger Kommunikation. Beiträge zu dem Workshop 'Methodologie und Datenanalyse'. Arbeiten zur Mehrsprachigkeit. Folge B, 11*, 2–25.
Rehbein, Jochen (2001). "Das Konzept der Diskursanalyse." In K. Brinker, G. Antos, W. Heinemann, & S. F. Sager (Eds.), *Text- und Gesprächslinguistik. Linguistics of Text and Conversation. Ein internationales Handbuch zeitgenössischer Forschung. An International Handbook of Contemporary Research. Bd.II* (pp. 927–945). Berlin and New York: de Gruyter.

Rehbein, Jochen, Kameyama, Shinichi, & Maleck, Ilona (1995). *Das reziproke Muster der Terminabsprache. Zur Modularität von Dialogen und Diskursen.* (*Verbmobil Memo 23.*) Hamburg: Germanisches Seminar.

Todd, Dundas A. & Fisher, Sue (1993). *The Social Organization of Doctor-Patient-Communication.* Norwood: Ablex.

Index

T